

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE VETERINARIA
Departamento de Producción Animal



**DIVERSIDAD GENÉTICA Y ESTABLECIMIENTO DE
PRIORIDADES EN ESQUEMAS DE CONSERVACIÓN:
EJEMPLO DE APLICACIÓN EN LA RAZA DE LIDIA**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR
PRESENTADA POR

David García Martínez

Bajo la dirección del doctor
Javier Cañon Ferreras

Madrid, 2008

- **ISBN: 978-84-692-4262-9**

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE VETERINARIA

Departamento de Producción Animal



**DIVERSIDAD GENÉTICA Y ESTABLECIMIENTO DE
PRIORIDADES EN ESQUEMAS DE CONSERVACIÓN.
EJEMPLO DE APLICACIÓN EN LA RAZA DE LIDIA**

Tesis Doctoral

David García Martínez

Madrid 2008

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE VETERINARIA

Departamento de Producción Animal



**DIVERSIDAD GENÉTICA Y ESTABLECIMIENTO
DE PRIORIDADES EN ESQUEMAS DE
CONSERVACIÓN. EJEMPLO DE APLICACIÓN EN
LA RAZA DE LIDIA**

MEMORIA PRESENTADA PARA OPTAR AL GRADO DE DOCTOR POR:

David García Martínez

Bajo la dirección del doctor:

D. Javier Cañón Ferreras
Madrid 2008

UNIVERSIDAD COMPLUTENSE DE MADRID
Facultad de Veterinaria



**DIVERSIDAD GENÉTICA Y ESTABLECIMIENTO
DE PRIORIDADES EN ESQUEMAS DE
CONSERVACIÓN. EJEMPLO DE APLICACIÓN EN
LA RAZA DE LIDIA**

Tesis Doctoral

David García Martínez
Madrid 2008

UNIVERSIDAD COMPLUTENSE DE MADRID

Facultad de Veterinaria

**DIVERSIDAD GENÉTICA Y ESTABLECIMIENTO
DE PRIORIDADES EN ESQUEMAS DE
CONSERVACIÓN. EJEMPLO DE APLICACIÓN EN
LA RAZA DE LIDIA**

*Memoria presentada por David García Martínez para optar al grado de Doctor,
realizada en el Departamento de Producción Animal de la Facultad de
Veterinaria de la Universidad Complutense de Madrid y dirigida por el Dr.
Javier Cañón Ferreras.*

Agradecimientos

En primer lugar, debo y quiero agradecerle a mi director, el Profesor Javier Cañón, no sólo su labor de dirección científica, sin la cual este trabajo que más o menos pretenciosamente da en llamarse tesis doctoral obviamente nunca habría sido posible, sino la confianza, académica, laboral y personal, que ha venido depositando en mí durante tantos años, muchas veces a pesar de mi propia incomprensión al respecto. Javier ve en mí algo que ni yo mismo veo, y los esfuerzos de toda índole que continuamente ha hecho por mí me hacen estar en deuda con él más allá de lo que los agradecimientos de una tesis puedan indicar.

La otra cabeza visible de este grupo de investigación al que pertenezco es la Profesora Susana Dunner. A ella le estoy igualmente agradecido por demostrar una paciencia más que meritoria tratando de hacerme entender los para mí oscuros mecanismos moleculares que están detrás de los números que en última instancia me llegan y con los que mi naturaleza y mi formación previa me impulsan a estar bastante más cómodo.

Continuando con el sector académico, mi agradecimiento ha de hacerse extensivo al Profesor Norberto Corral, por las ideas, el apoyo estadístico, y, sobre todo, por recordarme periódicamente ante mis constantes dudas, que en mi título de licenciado sigue diciendo que soy matemático. El apoyo moral recibido por su parte está a la altura, si no por encima, del puramente científico. Por tres cuartos de lo mismo debo darle las gracias al Profesor Carlos Carleos, quien además de con su amistad y su ánimo, me obsequia regularmente con su entusiasmo (no demasiado compartido) por las lenguas que se resisten a extinguirse y la informática variopinta.

Volviendo a casa, o lo que ha sido mi casa durante los últimos nueve años, el Laboratorio de Genética de la Facultad de Veterinaria, debo y quiero de nuevo mencionar a mis compañeros, y de nuevo por varios motivos. El primero es de una obviedad rotunda: sin ellos no habría tenido datos para realizar, ni esta tesis, ni ninguno de los análisis que han concluido en la serie de publicaciones a las que he contribuido en mayor o menor medida. Esa labor, a veces oscura, a veces desagradecida, de pasarse el día entre pipetas y aparatos que hacen cosas que no entiendo, está detrás de todo resultado que se produce, y por más que alguien

con mi formación siempre puede recurrir a inventarse datos en forma de simulaciones para poner a prueba las distintas metodologías, son siempre los datos reales los que en última instancia cuentan y engrandecen un trabajo o una publicación. Por todo ello, gracias, y gracias a todos.

Pero como no sólo de números vive el hombre, debo agradecerles más cosas. A los veteranos que aún siguen aquí con vinculaciones dispares debo estarles agradecidos por una amistad que trasciende las paredes del laboratorio. Juntos hemos compartido alegrías, tristezas, agobios, presiones, confidencias, risas y llantos, en innumerables jornadas diurnas en la facultad y en otras más numerables pero no menos memorables jornadas nocturnas en Madrid y más allá. Óscar es un tipo cuya clase, caballerosidad y savoir faire en general le preceden donde quiera que vaya, y su amistad y apoyo han sido una constante vía de escape para mí. Paloma y yo sabemos más el uno del otro de lo que la mayor parte de nuestros amigos más cercanos saben. La confianza y el entendimiento y apoyo mutuos que nos dedicamos son la prueba palpable de que no hace falta compartir orígenes, creencias ni ideologías para crear, mantener y estrechar lazos. Quiero cerrar este grupo con mis queridas Peruvians, Isabel y Susy, que mientras escribo estas líneas están pasando por momentos particularmente difíciles que me impulsan a utilizarlas (las líneas) para dejar patente el mayor de mis ánimos, que no es que sea muy grande dada mi naturaleza poco animosa, pero ahí va todo para ellas. Realmente he conocido en mi vida a muy pocas personas con un espíritu de trabajo y una dedicación como los suyos, y mi admiración sobrepasa mi gratitud hacia ellas, que de por sí no es poca (especialmente teniendo en cuenta que buena parte de mis ingresos en los últimos años se deben directamente al sudor de su frente). Yo no las entiendo muy bien, creo, y ellas a mí tampoco, pero nos queremos a nuestra manera.

No puedo olvidar al resto de compañeros con los que he compartido trabajo y vivencias a lo largo de estos años, los que siguen aquí y los que ya se han ido: Juan Pedro (los porteros argentinos no la saben bloquear, por más que te duela), Isaac (algún día Murphy se olvidará de ti, muy posiblemente para acordarse -más- de mí), Javi (las comidas no son lo mismo sin tí), Ana (por aquí seguiremos rescatando bichos, y les daré latas a los gatos en tu nombre), Natalia (haremos algo chulo con las perdices), Raquel (cuya pasión y vitalidad musical casi igualan a las mías, y cuya pasión y vitalidad científica superan a las mías con creces), Rodrigo (¡huevo!), Parra (al que aún no he visto de verde aceituna), Crisna (¿cómo lo hará para tener siempre una sonrisa en la boca que ofrecer desinteresadamente?), Bea (o La-Santa-con-la-que-compartí-piso-y-despacho-durante-años-e-incomprensiblemente-nunca-llegó-a-degollarme), Marisa (que seguro que está resultando tan gran mamá como compañera), Esther (que me soportó con estoicismo en el zaguán), Mariu (sigo recordando aquel encuentro sorprendente en La Vaguada), Ylenia (que después de todo este tiempo siempre me sigue felicitando las navidades y el cumpleaños pero soy tan impresentable como para olvidarme de responderle), Hosam (el único e incomparable Jodido), Luis (que debe ser de los pocos que en

lugar de irse de Asturias para buscar trabajo se va de Madrid hacia allí a encontrarlo, y con el que antes de que las cosas se torcieran nos echamos unas buenas risas), Ángel (otro con una capacidad de trabajo que yo nunca llegaré a alcanzar y que contribuyó con largas partidas de squash a mantener la línea que luego perdí tan lamentablemente), y alguno más que a mi memoria de pez se le escapa. Con todos ellos pasé buenos momentos personales y todos me han enseñado algo de esa cosa tan misteriosa que es la genética molecular.

El penúltimo, pero el mayor, de mis agradecimientos va dirigido a mis padres, por haberme querido, sabido e insistido en transmitir el valor que el respeto, la honradez y la educación, en todas las posibles acepciones del término, algunas (o muchas) de ellas tan vapuleadas hoy en día, han de tener en la vida de uno. La santa paciencia que han tenido, y que a día de hoy siguen teniendo, conmigo, es sólo la que un padre y una madre pueden tener. Cualquier otra persona me habría frito a sartenazos hace ya mucho tiempo. Intento estar a su altura, pero no puedo.

Finalmente, debo dejar constancia escrita de mi agradecimiento a Michi y a Bobo, por la (des)interesada compañía con la que me han obsequiado durante un buen puñado de años y que ha hecho mucho más llevadera mi estancia en este sitio tan curioso. Poca gente puede decir que ha trabajado con una gata ronroneando en el regazo mientras un gato se restriega cariñosa, que no sexualmente, contra su pierna, y cuando mira por la ventana puede contemplar una bucólica escena repleta de, ora cabritillos danzantes, ora burritos retozantes. Y (casi) ninguno de ellos caquético o hipopotasémico. Ah, ¡cuánto he aprendido y qué bien me lo he pasado!

Y ahora, al lío.

Índice

Resumen	iii
Introducción y objetivos	vii
Parte I.- Diversidad genética: conceptos, medida y contribuciones	1
1.- Concepto y necesidades de preservación de la biodiversidad	3
1.1.- Necesidad de conservación de la diversidad genética	6
1.2.- Amenazas y niveles de riesgo	10
1.3.- Estrategias de conservación	14
1.4.- Diversidad entre vs diversidad dentro. Heterosis vs. selección	16
2.- Medidas de diversidad genética	19
2.1.- Heterocigosis y estadísticos F	23
2.2.- Distancias genéticas	27
2.3.- Otras medidas clásicas de diversidad	39
2.4.- Coeficientes moleculares de parentesco	42
2.4.1.- Similitud y parentesco molecular	43
2.4.2.- Estimadores moleculares del parentesco genealógico	44
2.4.3.- Regresión log-lineal en la estimación de parentescos con información molecular	46
2.5.- Diversidad de Weitzman	52
2.5.1.- Definición y propiedades	53
2.5.2.- Formulación alternativa de D y representación gráfica	57
2.5.3.- Incorporación de información intrapoblacional	58
2.5.4.- Aproximación del algoritmo para un número grande de poblaciones	62
3.- Estrategias de asignación de prioridades en esquemas de conservación	71
3.1.- Contribuciones parciales, diversidad esperada, diversidad marginal y elasticidad de la diversidad	75
3.1.1.- Utilización de probabilidades de extinción genética en el cálculo de la diversidad marginal y elasticidad de la diversidad	78
3.1.1.1.- Variación en el tiempo de las frecuencias alélicas. Procesos de difusión	80
3.1.1.2.- Ecuaciones de Kolmogorov	82

3.1.1.3.- Probabilidad de extinción de un alelo al cabo de t generaciones	83
3.1.1.4.- Solución en poblaciones sometidas únicamente a deriva genética	84
3.2.- Minimización del parentesco medio y métodos Core Set	87
3.2.1.- Contribuciones al parentesco medio y descomposición en componentes entre y dentro de poblaciones	88
3.2.2.- Contribuciones óptimas para minimizar el parentesco medio. Método del Core Set	90
3.2.3.- Condición de mínimo de la solución del proceso de optimización	95
3.2.4.- Extensión del método Core Set para minimizar la varianza de un carácter cuantitativo	99
3.3.- Contribuciones globales combinadas	101
Parte II.- Conservación de la diversidad genética de la raza de lidia	103
4.- Introducción a la raza de lidia	105
4.1.- Origen de la raza de lidia	108
4.1.1.- Origen filogenético	108
4.1.2.- Origen histórico	110
4.2.- Estructura poblacional de la raza de lidia	113
5.- Material animal y molecular analizado	117
6.- Análisis, resultados y discusión	121
Parte III.- Conclusiones	141
Bibliografía	147
Apéndice: publicaciones relacionadas	161

Resumen

El trabajo expuesto en esta tesis profundiza en los aspectos matemáticos de la asignación de prioridades a un conjunto de unidades de conservación (poblaciones, líneas, razas, especies...) en el marco de programas de conservación de la diversidad genética. Los recursos a distribuir por parte de las administraciones son siempre limitados, y de ellos debe obtenerse el mayor rendimiento posible. En términos de diversidad, esto significa que deben repartirse proporcionalmente de acuerdo con la aportación que cada unidad haga a la diversidad global. El problema de la estimación de estas aportaciones no es trivial ni tiene una solución única. Conceptualmente existe además una disyuntiva sobre a qué dar más peso, si a las diferencias entre unidades o a la variabilidad dentro de ellas.

Se realiza un breve repaso inicial por las medidas tradicionales de diversidad genética, tanto entre como dentro de poblaciones, para a continuación adentrarse en dos concepciones recientes de la diversidad orientadas al posterior cálculo de diversos tipos de contribuciones a esa diversidad por parte de cada población y conseguir así un criterio para establecer prioridades. Los dos métodos estudiados, la diversidad de Weitzman y el Core Set, se analizan en profundidad, realizando aportaciones originales que contribuyen a consolidar sus propiedades en unos casos y a mejorar su comportamiento en otros. Se reseña también un tercer método de contribuciones ponderadas cuyo objetivo es tener un mayor control sobre el peso que los componentes *entre* y *dentro* de la diversidad ejercen en la decisión final.

La segunda parte de esta tesis está dedicada a ejemplificar la metodología anterior en un caso práctico y singular como es el de la raza de lidia, en el que las unidades de estudio no son razas dentro de una especie,

como es habitual en los trabajos sobre diversidad, sino linajes, llamados en este caso *encastes*, dentro de una única raza. Se estudian los diversos parámetros relacionados con la diversidad y se llevan a cabo los distintos tipos de ponderaciones estudiados en la parte primera, analizando y comentando los resultados obtenidos.

Esta tesis es, por tanto, un trabajo fundamentalmente metodológico, de revisión en profundidad de dos formas distintas de abordar un mismo problema, con algunas aportaciones propias e igualmente metodológicas a ambas, y un ejemplo práctico en el que se describe cómo se aplicarían en un trabajo de campo.

Abstract

The work in this thesis explores the mathematical aspects of priority settings in a set of conservation units (populations, lines, breeds, species...) in the framework of genetic diversity conservation programmes. Resources from governments and public organizations are always limited, and maximum profit must be obtained from them. In terms of diversity this means that they must be split proportionally, according to the share each unit has in the global diversity. However, the estimation of these contributions is not trivial nor has a unique solution. There is also a conceptual dichotomy regarding the importance of between units, as opposed to within units, variation.

Initially, an outlook on the traditional between and within-population diversity measures is performed, to move then towards deepening into two recent criteria for genetic diversity measurement. Both of these new concepts were developed to be incorporated in the calculation of different types of unit contributions and the establishment of priorities among those units. The two methods, Weitzman's and the Core Set method, are studied in depth, adding original contributions that help establishing their properties and improving their performance. Also reviewed is a third method of weighted contributions, whose

objective is to have more control over the weights that the between- and within-components have in the final decision.

The second part of this thesis is devoted to exemplify the previous methodology in a practical and singular case, as is the lidia breed. Unlike most genetic diversity studies, the units under study here are not breeds within a species, but lineages (called *encastes*) within a breed. Several diversity-related parameters are studied, and the different types of contributions defined previously are analysed and commented.

This thesis is therefore basically a methodological work, of in-depth review of two different ways of approaching one same problem, with some original methodological contributions to both, and a real data example describing how to apply them in a field work.

Introducción y objetivos

El desarrollo sostenible, la conservación de la biodiversidad, el respeto al medioambiente o el cambio climático son aspectos que de forma cada vez más habitual y con intensidad creciente traspasan la frontera de la investigación y divulgación científica para convertirse en asuntos de debate en la calle y las instituciones políticas, económicas y sociales. Resulta frecuente encontrar artículos en la prensa diaria sobre resoluciones políticas o recomendaciones de organismos oficiales encaminadas a frenar el deterioro medioambiental en cualquiera de sus vertientes: recursos hídricos, contaminación atmosférica, deforestación, pérdida de biodiversidad...

En el primer cuarto de este año 2008, sin ir más lejos, se han producido en nuestro país importantes controversias sobre la Ley de Biodiversidad aprobada por el Gobierno Español a finales de 2007, y el cambio climático ha ocupado buena parte del debate político previo a las elecciones generales de Marzo.

Lo ecológico está, pues, de moda. Pero como casi siempre sucede cuando algo de índole científica acaba por calar en la mente colectiva de nuestra sociedad, la ciencia lleva ya años preocupada y dedicando recursos ingentes a tratar y estudiar el problema. Por ejemplo, sólo en el ámbito de la diversidad genética en animales de granja evaluada con marcadores moleculares, que a la postre es el objeto de esta tesis doctoral, se publicaron en el período de 1996 a 2003, más de cincuenta estudios en revistas científicas de impacto, y en ese mismo año 2003 había cerca de 90 proyectos de investigación recién finalizados o aún en curso relacionados con la materia que recababan muestras de razas otros tantos países a nivel mundial.

Pero quizás el acontecimiento relevante más reciente relacionado con la preservación de la biodiversidad haya sido la inauguración en febrero de 2008 del banco global de semillas de Svalbard, un complejo de cámaras excavadas en la roca y el hielo de una isla mil kilómetros al norte de la costa Noruega destinado a conservar semillas de prácticamente todas las especies y variedades cultivadas por el hombre como reserva mundial de emergencia en previsión de catástrofes. Hasta 4.5 millones de muestras se podrán llegar a almacenar en unas condiciones ideales, constuyendo así un reservorio de recursos sin parangón en la historia de la humanidad.

La creación de “copias de seguridad” para hacer frente a poblaciones diezmadas por epidemias, catástrofes naturales o eventos relacionados con la actuación del hombre es una de las diversas aplicaciones de los programas de conservación de la diversidad. Las funciones de estos programas van, sin embargo, mucho más allá, ya que pueden contribuir a la revitalización de zonas rurales, a mantener el nivel de las actividades productivas asociadas a determinadas razas y, en general, a evitar que una raza o una especie se consuma víctima de una espiral de extinción.

A lo largo de todo el mundo se están llevando a cabo diseños y puestas en práctica de esquemas de conservación de recursos zoo y fitogenéticos. Se aplican esquemas de todo tipo, in-situ, ex-situ o in vitro, dependiendo de las necesidades concretas y los objetivos que se persigan, y se priorizan las contribuciones de las distintas especies, de las distintas razas dentro de una única especie o de las distintas líneas dentro de una única raza, de forma que se optimice la diversidad total conservada. Y es que la existencia de recursos limitados impone la necesidad de establecer un criterio según el cual se dé mayor prioridad en su asignación a aquellas unidades que más diversidad aporten al conjunto. Los fondos, obviamente finitos, que un organismo internacional destine a la conservación de una especie han de distribuirse entre las distintas razas que la forman, y esa distribución ha de hacerse de forma óptima. Como óptima debe ser la manera en que se elijan las unidades a entrar en un banco de germoplasma, que tendrá una capacidad determinada. El problema es que no existe un criterio único para tomar estas decisiones, y de

hecho ni siquiera hay un consenso sobre el peso que los distintos componentes de la diversidad deben tener a la hora de decidir la distribución de recursos entre las distintas unidades. Tampoco resulta sencillo discernir cómo afecta en la práctica cada unidad económica invertida en una unidad de conservación en el aumento de su diversidad, o en su caso, en el freno de la pérdida de ella.

Tanto la FAO a nivel mundial como la EAAP en el entorno europeo publican periódicamente circulares con directrices a tener en cuenta para la conservación de los recursos genéticos animales, con especial énfasis en la sostenibilidad de las razas autóctonas. Cada vez existe mayor conciencia sobre la importancia de estas razas, como garantes por un lado de las aptitudes adaptativas a los entornos concretos en los que subsisten, y por otro como reservorio de una diversidad genética potencialmente utilizable en caso de necesidad, además del valor histórico o cultural que pueden representar para un determinado pueblo o región, que en la práctica se puede traducir en valor añadido del desarrollo sostenible de las zonas rurales. Es necesario, pues, aplicar políticas que fomenten la conservación de la diversidad genética, más aún en casos en los que la viabilidad de la raza o de un conjunto de razas se ve amenazada por problemas de consanguinidad derivados del aislamiento reproductivo, la deriva genética y la ausencia de control en los apareamientos.

En esta tesis doctoral se pretende, como objetivo general, dar respuesta a la pregunta de cómo medir esa diversidad genética, pero sobre todo cómo establecer prioridades de forma que se conserve la mayor cantidad posible. Se estudiarán los distintos criterios existentes y se intentará mejorar metodológicamente alguno de ellos para que tenga en cuenta un mayor número de factores. Como ejemplo de aplicación, se estudiará la diversidad genética en una raza con una serie de particularidades ausentes en la mayoría de razas ganaderas, como es la raza de lidia, estableciendo la importancia relativa de cada encaste analizado según los distintos criterios existentes.

El contenido de esta tesis estará, pues, estructurado en dos secciones. La primera de ellas tratará sobre la diversidad genética como concepto. Se hará un breve recorrido por la situación actual de la diversidad genética animal,

se plantearán distintas estrategias globales de conservación, se discutirán las ventajas e inconvenientes de aplicar programas que favorezcan la diversidad dentro de unidades en contraposición a aquéllos que dan más importancia a la diversidad entre unidades, se introducirán distintas medidas clásicas de diversidad y se profundizará en los avances metodológicos más recientes en lo que a priorización de unidades en función de su aportación a la diversidad conjunta se refiere. En la segunda sección se introducirá la peculiar idiosincrasia de la raza de lidia, se describirá su situación actual en términos de diversidad genética y se hará un estudio de su conservación.

PARTE I

Diversidad genética: conceptos, medida y contribuciones

1.- Concepto y necesidades de preservación de la biodiversidad

De acuerdo con la FAO, la Organización para la Alimentación y la Agricultura de la ONU, se define la biodiversidad, o diversidad biológica, como “la variación de la vida en todas sus formas, niveles y combinaciones, incluyendo la diversidad genética, la diversidad en las especies y la diversidad en los ecosistemas” (FAO, 1999). Se puede por tanto estudiar la diversidad en todos los niveles de un ecosistema: la de especies dentro de ecosistemas, la de razas o poblaciones dentro de especies, o la existente dentro de esas poblaciones, entendiéndose la diversidad global como la combinación de todas esas fuentes de variabilidad. En particular, se puede estudiar la diversidad genética como parte fundamental de la biodiversidad, y al igual que ésta, se puede descomponer en niveles: ecosistema, especie, raza, línea, rebaño, individuo...

Si en las especies salvajes existe un interés por preservar esa biodiversidad debido fundamentalmente a su valor biológico y a la sostenibilidad de los ecosistemas, a las especies domésticas se une el valor añadido de su interés para el ser humano. Actualmente casi dos mil millones de personas dependen al menos parcialmente de animales domésticos para su supervivencia, y un 12% de ellas lo hace casi completamente (FAO, 1999). A la diversidad biológica del mundo agrario se la conoce como agrobiodiversidad, y es el componente de la biodiversidad global que contribuye a la producción de alimentos y de productos agrarios.

El ser humano lleva al menos 12000 años involucrado en la domesticación o el uso de plantas y animales salvajes, pese a lo cual se le ha dado poco uso a la diversidad animal que se encuentra en el planeta. De 50000 especies aviarias y de mamíferos conocidas, menos de 30 han tenido un uso extendido para la agricultura. De ellas, menos de 14 acaparan el 90% de la producción ganadera conjunta (FAO, 1999).

Por otra parte, la diversidad genética de las especies domesticadas se ha utilizado de forma extremadamente eficaz. A lo largo de varios siglos, los granjeros y criadores han seleccionado animales para una amplia variedad de características, resultando en el desarrollo de las 4500-5000 razas actuales. Sólo nueve de las 14 especies más importantes (vaca, caballo, asno, cerdo, oveja, búfalo, cabra, gallina y pato) pueden reunir nada menos que 4000 razas en todo el mundo (FAO, 1999).

Aún así, el hecho es que hoy en día alrededor del 35% de las razas mundiales están clasificadas como razas con alto riesgo de extinción, y en promedio cada semana desaparecen permanentemente dos de ellas (Scherf, 2000), perdiéndose así para siempre una importante fuente de recursos genéticos.

1.1.- Necesidad de conservación de la diversidad genética

El ganado se ha convertido en un importante elemento cultural, esencial en el mantenimiento de muchos estilos de vida tradicionales. Pese a que la contribución de los animales domésticos a la agricultura y al progreso económico global no ha sido adecuadamente evaluada ni apreciada, las razas autóctonas constituyen en sí mismas una fuente de valores que hacen de su conservación una necesidad prioritaria, y lo cierto es que en los últimos tiempos se está produciendo un aumento en la conciencia internacional del papel esencial que juegan en la producción y la productividad de la agricultura los recursos genéticos animales y un consiguiente aumento en la cantidad de estudios sobre diversidad genética en animales domésticos (Amos y Balmford, 2001; Barker, 2001; INRA y CIRAD, 2002; Baumung *et al.*, 2004; Ollivier y Foulley, 2004; Bennewitz y Meuwissen, 2005).

En el año 2003 la FAO distribuyó un cuestionario a los distintos grupos de investigación involucrados en estudios de diversidad genética. Los resultados de este cuestionario, así como de una revisión de la bibliografía relacionada, con el objetivo de esbozar una perspectiva de la situación de este tipo de estudios en términos de razas, tamaños muestrales, información molecular y metodología aplicada, se pueden encontrar en Baumung *et al.* (2004). Se exhiben datos procedentes de 87 proyectos de investigación, finalizados o en curso, con muestras procedentes de 93 países distintos, estudiándose una media de 18 razas por proyecto, hasta un máximo de 120, aunque en bastantes casos (12%) el proyecto estaba centrado en una única raza. La investigación en el campo de la diversidad genética es, por tanto, abundante, y el volumen de información que genera no lo es menos. Baste

decir que en el momento de la encuesta de la FAO, 45 proyectos estaban aún en curso, pero ya se habían publicado 47 artículos relacionados con ellos.

Además del valor económico asociado intrínsecamente a la producción ganadera, existen una serie de valores socioeconómicos y culturales a tener en cuenta para conservar una raza. Esto se da especialmente en razas locales, autóctonas de regiones muy concretas y adaptadas perfectamente a su entorno a través de décadas e incluso siglos de domesticación y selección. No suelen ser razas de gran rendimiento productivo, y su adaptación extrema a las condiciones locales hace en ocasiones difícil su expansión o exportación a otros lugares. Así, en el plano cultural se da con cierta frecuencia una identificación especial de los habitantes de una determinada zona con sus razas locales, a través del arte, el folklore o el propio sentimiento de la raza como algo propio. Las razas autóctonas se convierten de esta forma no sólo en señal identitaria de la propia zona, sino en parte integrante del paisaje, lo que les otorga a su vez un valor ecológico por una parte, y socioeconómico por otra, proveniente de la interacción de su presencia con el conjunto de paisaje, naturaleza y gestión de las granjas. El turismo rural se ve atraído por el entorno del que forman parte los animales, y las granjas especializadas en las razas locales dan salida a productos con una imagen ecológica y altamente apreciados por un determinado sector de consumidores.

La diversidad genética permite al ganado al reto de nuevas enfermedades, infecciosas o parasitarias, amplias variaciones en la disponibilidad y calidad de la comida y el agua y otros factores limitantes. Las razas adaptadas a un entorno local pueden tener unas capacidades productivas modestas en comparación con razas altamente especializadas, pero en cambio son muy eficientes en el uso de los recursos disponibles, y son sostenibles a largo plazo. De hecho, en muchas ocasiones, animales importados provenientes de razas muy extendidas de mayor productividad no se han reproducido o sobrevivido tan bien como las razas adaptadas localmente (FAO, 1999).

La diversidad genética es, pues, un valor en sí mismo, y su mantenimiento es un objetivo prioritario en los programas de conservación de biodiversidad (Fernández *et al.*, 2004). Es un valor a tener muy en cuenta puesto que condiciona muchos otros. Poseer unos recursos genéticos suficientemente amplios garantiza la adaptación y viabilidad de una especie o una raza a entornos de producción variables (Barker, 1999; INRA y CIRAD, 2002; Gandini y Oldenbroek, 2007). Desde arbitrariedades tales como cambios en los gustos del consumidor que exijan alterar los objetivos de selección, hasta la aparición de nuevas enfermedades, pasando por distintos tipos de alimentación o manejo, la conservación de la diversidad genética es una salvaguarda frente a cualquier reto adaptativo proveniente de alteraciones en las circunstancias de producción.

Como salvaguarda actúa también en casos de diezmas repentinas en el censo de una determinada especie, debido a catástrofes naturales, plagas o guerras. Los acontecimientos vividos en los últimos tiempos, en forma de epidemias y amenazas de armas biológicas, han hecho crecer en los gobiernos la conciencia sobre el valor que la conservación de los recursos genéticos actuales puede llegar a tener en este sentido. Una “copia de seguridad” en forma de banco de germoplasma es una forma rápida y razonablemente barata de poder recuperar una raza en caso de llegar al borde de la extinción, y una garantía para el mantenimiento de los sistemas de producción de alimentos a escala nacional.

La diversidad genética ofrece además una oportunidad única para la investigación científica. De no existir suficiente variabilidad no se podrían llevar a cabo las distintas líneas de investigación orientadas a la identificación de genes influyentes en caracteres productivos, reproductivos o de calidad de los distintos productos, o bien relacionados con la resistencia o susceptibilidad a determinadas enfermedades. La existencia de tipos muy distintos de razas y la posibilidad de realizar cruces orientados permite disponer de unos niveles de heterocigosis, desequilibrio de ligamiento y diversidad alélica imprescindibles para detectar posibles asociaciones de marcadores moleculares de localización conocida con esos genes, o dicho de otra forma, para poder señalar con

precisión zonas en el genoma que influyen en los caracteres de interés (Oldenbroek, 2007).

La pérdida de variabilidad genética produce una pérdida de todos estos activos, pero no sólo eso, sino que el empobrecimiento genético puede en última instancia llevar a una raza o una especie a la extinción más allá de la influencia de agentes externos. Una mala política de conservación de la diversidad genética, o en el peor de los casos, la ausencia de ella, se traduciría en la práctica en un progresivo deterioro de la raza. Cuando se trata con poblaciones de tamaño reducido y no se lleva a cabo una estrategia de apareamientos dirigidos a obtener mínima consanguinidad o mínimo parentesco, se corre el riesgo de aumentar los niveles de endogamia, lo que se traduce en depresión consanguínea (infertilidad, susceptibilidad a enfermedades, etc.). Ésta a su vez provoca que haya un descenso en los niveles reproductivos y disminuya la tasa de crecimiento de la población, lo que cierra el círculo haciendo que el censo caiga aún más, en lo que se conoce como espiral de extinción (Gilpin y Soulé, 1986).

1.2.- Amenazas y niveles de riesgo

Además de la mala gestión de los recursos genéticos, existen factores exógenos que también influyen en su deterioro. Actualmente las amenazas a los recursos genéticos agrícolas provienen de varios frentes (Oldenbroek, 2007):

- Cambios socioeconómicos, urbanización y factores políticos que llevan a una intensificación de la producción y la consiguiente eliminación de razas autóctonas o de creación reciente y aptitud múltiple.
- Comercialización global de material genético exótico, acompañado de la sustitución de razas locales por exóticas.
- Liberalización de mercados para productos animales que perjudican y ralentizan el desarrollo de sistemas de producción locales con razas autóctonas.
- Pérdida de diversidad cultural y hábitats y estilos de vida tradicionales.
- Cambios en los ecosistemas que requieran distintas capacidades adaptativas en los animales involucrados.
- Guerras, inestabilidades políticas, enfermedades y desastres naturales, como fuentes de destrucción no sólo de las razas en sí mismas, sino a menudo de las infraestructuras de producción ganadera.

Así las cosas, los distintos organismos internacionales han propuesto una clasificación del nivel de riesgo de las distintas razas, atendiendo a criterios como el censo, el número de reproductores o el incremento en endogamia.

La FAO establece las siguientes categorías de riesgo: extinta, crítica, crítica-mantenida, en peligro, en peligro-mantenida y fuera de peligro (Scherf, 2000). Se basa para la clasificación de una raza en una u otra categoría en el número de reproductores machos, hembras y totales. Así, una raza estará:

- *Extinta* si no existen reproductores o reproductoras. Es imposible recuperarla.
- En estado *crítico* si hay menos de 100 reproductoras, menos de 5 reproductores, o el número reproductores totales está en descenso y por debajo 120, y menos de un 80% de ellos son de pura raza.
- *Crítica-mantenida* si está en estado crítico pero existe un programa de conservación o de mejora para la raza.
- *En peligro* si hay menos de 1000 reproductoras, o bien menos de 20 reproductores, o bien el número de reproductores totales se encuentra en ascenso y entre 80 y 100, con más del 80% de ellos de pura raza, o bien esta cifra está descendiendo desde unos valores de entre 1000 y 1200 y menos de un 80% de ellos son de pura raza.
- *En peligro-mantenida* si está en peligro pero existe un programa de conservación o de mejora para la raza.
- *Fuera de peligro* si el número de reproductoras es superior a 1000, el de reproductores superior a 20 o el número total de reproductores está por encima de 1200 y en aumento.

La EAAP (Asociación Europea de Producción Animal) en cambio, define unos criterios basados en la tasa de consanguinidad en 50 años, y clasifica a las razas como en peligro crítico, en peligro, ligeramente en peligro, posiblemente en peligro y fuera de peligro (EAAP, 1998). Los incrementos en consanguinidad correspondientes a cada una de las categorías son los siguientes:

- *En peligro crítico*: ΔF superior al 40%
- *En peligro*: entre un 26 y un 40%
- *Ligeramente en peligro*: entre un 16 y un 25%
- *Posiblemente en peligro*: entre un 5 y un 15%
- *Fuera de peligro*: por debajo del 5%

Como criterios adicionales para ajustar la categoría de riesgo entre dos adyacentes, la EAAP contempla la proporción de registros en el libro genealógico, el cambio en el número de reproductores, el porcentaje de animales de pura raza y de inmigración y el número de rebaños.

Finalmente, la Unión Europea, a través de su Reglamento CE de la Comisión 445/2002, determina para cada especie subvencionable un número mínimo de cabezas cruzadas en pureza por debajo del cual se puede considerar a una raza en riesgo de abandono. Esta cifra se situaría para bovino en 7500, para ovino y caprino en 10000, para equino en 5000, para porcino en 15000, y en 25000 para especies avícolas.

El riesgo de desaparición de una raza puede ser utilizado como criterio de priorización en estrategias de conservación de la diversidad, pero no es trivial proporcionar un valor objetivo del riesgo real de extinción de una raza. No sólo los grandes organismos internacionales no comparten criterios comunes para categorizarlo, sino que no resulta difícil pensar en todos los factores que pueden llegar a influir en el grado de peligro que pueda tener una raza, y cómo dos razas que aparentemente pudieran estar en la misma categoría de acuerdo con la FAO, la EAAP o el reglamento de la UE, pueden estar en la práctica sometidas a riesgos muy distintos. Factores sanitarios, sociales, políticos, económicos o culturales, y por supuesto, genéticos, alteran notablemente las posibilidades de supervivencia de una población.

Como se verá más adelante, otras estrategias buscan el mantenimiento de la máxima diversidad posible, pese a que tampoco hay un consenso universal en cuanto a cómo medir y a qué niveles (especie, raza, rebaño, individuo) considerar esa diversidad; incluso se pueden plantear estrategias de

utilidad máxima, en las que lo que se trata de optimizar es el rendimiento económico de los recursos invertidos en la conservación (Reist-Marti *et al*, 2006). Tampoco en este caso es fácil proporcionar funciones objetivo que relacionen claramente unidades monetarias invertidas con beneficios obtenidos. Otras estrategias sugieren combinar la diversidad con la utilidad o con el mérito para determinados caracteres y la viabilidad de las poblaciones (Weitzman, 1998; Barker, 2001; Piyasatian y Kinghorn, 2003)

1.3.- Estrategias de conservación

Cuando se plantea un programa de conservación de la diversidad existen varias formas de llevarlo a cabo en la práctica, dependiendo de si el material genético se conserva en forma de gametos o embriones congelados o en los propios animales vivos, y de si éstos se mantienen en su hábitat natural o se crían fuera de él.

El caso más frecuente es aquél en el que se conservan los animales vivos y en su medio natural o en el que habitualmente producen. Es la llamada conservación “in situ”. Es preferible, ya que los objetivos de conservación se alcanzan de forma óptima y ofrece las mejores posibilidades de utilización del ganado. Además, el desarrollo de la raza continúa y facilita la adaptación a circunstancias cambiantes. El problema de este tipo de conservación proviene del riesgo de consanguinidad y efectos de deriva genética presentes en el mantenimiento de poblaciones con un censo reducido, como suelen ser las que se someten a estos programas (Gandini y Oldenbroek, 2007). Aún así, es la estrategia recomendada por la FAO para un uso sostenible de zonas rurales (FAO, 1999).

La conservación “ex situ” es aquella que se lleva a cabo fuera del entorno natural del animal. Puede ser “in vivo ex situ” o “in vitro”. La conservación “in vivo” mantiene animales vivos fuera de las condiciones normales de manejo, o fuera de la zona en la que se formó la población o en la que se encuentra en ese momento, como zoológicos o reservas animales. El coste de este tipo de conservación es menor, pero mantener la raza fuera de su entorno condiciona su adaptabilidad futura a cambios en él, además de no proporcionarle ningún beneficio en términos de sostenibilidad y desarrollo rural, y en cualquier caso la raza sigue expuesta a efectos como la deriva genética.

En muchas ocasiones la diferencia entre “in situ” y “ex situ” es difícil de determinar, por lo que se tiende a agrupar estas dos estrategias como “in vivo”, en contraposición a la conservación “in vitro” o crioconservación. En este caso no se conservan animales vivos, sino que se almacenan gametos (óvulos o esperma) o embriones congelados en nitrógeno líquido. Aunque las técnicas de conservación de material genético están suficientemente asentadas, aún hay técnicas como la transferencia nuclear que necesitan ser perfeccionadas para garantizar que se pueda regenerar una raza después de un almacenaje de este tipo (Oldenbroek, 2007). Aún así, esta técnica es óptima cuando lo que se pretende es tener una reserva de recursos genéticos en previsión de catástrofes que amenacen la supervivencia de la raza o la extingan de facto. Presenta además la ventaja de no estar expuesta a efectos como la deriva, aunque por otra parte la población “reconstruida” no habrá experimentado ninguna adaptación a un entorno potencialmente distinto de aquél en el que la raza habitaba en el momento de la selección del material genético para la crioconservación.

1.4.- Diversidad entre vs diversidad dentro. Heterosis vs. selección

Cuando se analiza la variabilidad en un conjunto de elementos, hay dos fuentes principales que contribuyen a ella: las diferencias entre elementos y la propia variabilidad existente dentro de cada uno de ellos. Existen diversas formas de medir cada uno de esos componentes, y, lo que es más importante, distintos puntos de vista sobre qué valorar más y cómo combinar ambos tipos de información. Dar preferencia a la diversidad entre o a la diversidad dentro a la hora de establecer prioridades de conservación es una decisión difícil de tomar. Ambos componentes deberían ser tenidos en cuenta, ya que aproximadamente un 50% de la variabilidad genética total en una especie se debe a las diferencias entre razas (Barker, 2001; INRA y CIRAD, 2002; Wooliams y Toro, 2007), procediendo la otra mitad de las diferencias entre animales dentro de cada raza. El peso que se le debe dar a cada uno de estos elementos, sin embargo, no es un asunto trivial.

Es frecuente que los estudios de diversidad en un conjunto de razas se lleven a cabo considerando especialmente las diferencias entre ellas, medidas a través de distancias genéticas (Barker, 1999; Ollivier y Foulley, 2002; Baumung *et al.*, 2004), aunque los procedimientos que únicamente consideran la variabilidad entre unidades, especialmente cuando éstas son razas y no especies, y el objetivo es determinar prioridades de conservación, han sido criticados, y se sugiere que la diversidad dentro de las razas sea tenida en cuenta (Barker, 2001; Caballero y Toro, 2002; Eding *et al.*, 2002; Bruford, 2004). En cuanto a utilidad práctica, favorecer la diversidad dentro de razas es útil en términos de selección y adaptación, mientras que darle más peso a la diversidad entre razas permite optimizar cruzamientos y crear poblaciones híbridas con el fin de explotar por selección la mayor diversidad generada. Las

poblaciones endogámicas con poca variabilidad son menos viables que poblaciones grandes con tamaños efectivos elevados. Por otra parte, combinar poblaciones alejadas entre sí, portadoras de distintos alelos segregando a frecuencias elevadas (y por tanto con un exceso de homocigotos) produce nuevas poblaciones con una alta variabilidad debido a los nuevos heterocigotos.

El problema es que hay una gran paradoja subyacente a la discusión “entre vs. dentro”. Si uno tiende a favorecer la diversidad entre poblaciones, las unidades más alejadas aparecen como prioritarias; en el caso de animales domésticos, las razas más alejadas son aquéllas que con más intensidad han sufrido procesos de deriva y selección, de forma que suelen presentar altos niveles de endogamia. Por tanto, son básicamente meros depósitos de variantes alélicas que raramente se encuentran en otras razas, pero no aportan diversidad de forma intrínseca, y habría razones para argumentar que se debería dar prioridad a razas con tamaños efectivos elevados en lugar de a estas razas genéticamente agotadas. Pero por otro lado, ¿qué sentido tiene invertir más recursos en razas o poblaciones que se encuentran de por sí fuera de todo peligro gracias a su estructura genética saneada?

Los enfoques teóricos utilizados generalmente para determinar prioridades de conservación están bien fundados y contruidos, y las metodologías se ajustan perfectamente a los objetivos propuestos, pero no se ajustan tanto a las demandas reales. Si los recursos se distribuyeran de la forma en la que los potenciales de conservación o las contribuciones parciales a la diversidad de Weitzman (Weitzman, 1992, 1993, ver secciones 2.5 y 3.1) aconsejan, la pérdida de diversidad entre poblaciones sería ciertamente minimizada en teoría. Pero la única manera real de preservar esa diversidad entre poblaciones sería mantenerlas tan alejadas como fuera posible, es decir, mantener o incrementar la presión de selección y el efecto de deriva, dando como resultado poblaciones genéticamente degradadas. Esta estrategia maximizaría la diversidad en hipotéticos cruzamientos, pero en primer lugar, en la práctica ninguna organización asignaría fondos para incrementar la consanguinidad de una población, y en segundo, es poco probable que los

criadores y ganaderos de razas locales aceptaran una política orientada a forzar cruzamientos entre las distintas razas. Además, es más favorable un relativo intercambio genético entre ellas que una política de aislamiento total (Pääbo, 2000).

Por otro lado, los métodos enfocados a la minimización del parentesco (Caballero y Toro, 2002; Eding *et al.*, 2002) resultan en prioridades opuestas a las sugeridas por los métodos basados en distancias, favoreciendo por tanto a las razas menos amenazadas, aquéllas que en la práctica menos ayuda necesitan. El razonamiento subyacente a estas metodologías es igualmente poco práctico, ya que se basa en la idea de crear una población sintética en la que el parentesco resulta minimizado, y se buscan las contribuciones a esa metapoblación que alcancen esa condición de mínimo. Esto es útil en programas de crioconservación dirigidos a crear una reserva de recursos genéticos en prevención de desastres naturales o provocados por el ser humano que llevaran a la especie al límite de la extinción y ésta tuviera que ser reintroducida. Pero para programas de conservación *in situ* vuelve a crear un conflicto, siendo como es poco probable que los criadores aceptaran la creación de una nueva metapoblación mixta.

La realidad también juega en contra de cualquiera de estos enfoques, en la forma de las políticas de concesión de fondos actualmente en vigor. La UE, por ejemplo, prevé que se asignen fondos por igual a todas las razas incluidas en programas de conservación, con independencia de censos, tamaños efectivos, niveles de peligro o prioridades publicadas en la bibliografía.

Teniendo en cuenta todo esto, podría ser interesante moverse a un nivel inferior en la escala taxonómica y determinar prioridades de conservación dentro de razas, es decir, al nivel de rebaños, en lugar de (o complementariamente a) prioridades al nivel de raza. La implementación práctica de estas metodologías es más accesible cuando la hipótesis de partida se reduce a cruzar rebaños en lugar de enfrentarse a un planteamiento de cruzamientos entre razas.

2.- Medidas de diversidad genética

Desafortunadamente no existe una única definición de diversidad genética. Son numerosos los parámetros, indicadores y medidas que se pueden entender como tal. En esta sección se revisarán brevemente las formas clásicas de medir la diversidad genética, para profundizar a continuación en dos medidas más elaboradas, que se han utilizado ampliamente en la bibliografía más reciente, especialmente con el objetivo de establecer prioridades entre distintas razas de una misma especie según sus contribuciones a la diversidad conjunta.

Una primera forma de entender la diversidad genética es a través de la variabilidad fenotípica. Utilizando el concepto de varianza estadística, la diversidad genética sería equiparable a la heredabilidad, o la proporción de la varianza de un carácter atribuible a causas genéticas (Woolliams y Toro, 2007). Esta concepción presenta algunos inconvenientes, ya que no siempre es posible o, en el mejor de los casos, sencillo, calcular los componentes de varianza correspondientes, y no es una medida fácilmente comparable, ya que depende del ambiente, que no suele ser uniforme para las distintas razas. Además, distintos fenotipos o caracteres presentan distintas heredabilidades, lo que cuestiona aún más su utilidad operativa.

Otra serie de medidas frecuentemente utilizadas como indicadores de la diversidad de una población vienen dadas mediante la información genealógica. Utilizando los pedigrís se pueden calcular parámetros como el tamaño efectivo, el nivel y el incremento por generación de endogamia, coeficientes de parentesco... Así, cuanto más emparentados estén los individuos, mayores sean los niveles de endogamia, o menores sean los tamaños efectivos, y menor será la variabilidad genética presente en la raza. En la práctica no siempre se tiene una información detallada de las genealogías, con lo que el cálculo de este tipo de parámetros no es factible, teniendo que recurrirse a marcadores de ADN, bien para obtener estimaciones

moleculares de sus valores, bien para computar otros tipos de diversidad genética.

Los marcadores moleculares juegan hoy en día un papel crucial en los estudios de diversidad genética (Durand-Tardif *et al.*, 2000; Baumung *et al.*, 2004; Bruford, 2004). Los más utilizados actualmente en estos estudios son generalmente de tipo microsatélite o SNP -Polimorfismo de Nucleótido Simple- (Baumung *et al.*, 2004), y dependiendo de su localización en el genoma miden diversidad neutra o adaptativa. Cuando un marcador se encuentra localizado en una zona no codificante del genoma, es decir, una zona de ADN que no se transforma (codifica) en proteínas, y por tanto no tiene ningún efecto real en el individuo, puede ser utilizado para estudiar cómo ha evolucionado la variabilidad neutra, aquélla que no se ve afectada por selección natural o artificial, sino únicamente por efectos poblacionales como la deriva genética o los cuellos de botella. Si, en cambio, el marcador se sitúa en la parte codificante de un gen, la que tiene influencia en algún fenotipo, una selección para alguna característica relacionada con ese fenotipo tendrá un efecto detectable en el comportamiento del marcador, de ahí que estos marcadores se utilicen para medir diversidad adaptativa (Bamshad y Wooding, 2003; Wiener *et al.*, 2003; Akey *et al.*, 2004; Bonin *et al.*, 2007).

Existe cierta controversia sobre la importancia relativa de la diversidad neutral frente a la adaptativa en los planes de conservación (Moritz, 1994; Bowen, 1999; Crandall *et al.*, 2000; Fraser y Bernatchez, 2001). Tradicionalmente se ha evaluado la diversidad adaptativa a través del estudio de caracteres cuantitativos, por ser éstos los objetivos principales sobre los que actúa la selección natural (Falconer y Mackay, 1996; McKay y Latta, 2002), y asumiendo que la variabilidad observada en estos caracteres refleja el potencial adaptativo de la población (Lynch, 1996; van Tienderen *et al.*, 2002). Estimar esta variabilidad es costoso, y la accesibilidad, coste y rapidez de los métodos moleculares utilizados para medir la diversidad neutral con marcadores de ADN han hecho de éstos la opción más habitual para medir la diversidad en términos globales (Petit *et al.*, 1998; Frankham *et al.*, 2002; Garner *et al.*, 2005). Sin embargo, recientes avances en las técnicas

moleculares y estadísticas han permitido disponer de herramientas para identificar regiones genómicas sometidas a selección en una determinada población (Black *et al.*, 2001; Luikart *et al.*, 2003; Beaumont y Balding, 2004), y se han propuesto índices basados en información sobre la variabilidad adaptativa a nivel molecular para establecer prioridades de conservación entre poblaciones (Bonin *et al.*, 2007).

Tradicionalmente los marcadores moleculares se han utilizado para calcular una serie de parámetros relacionados con la diversidad, tanto la existente dentro de una raza como la imputable a las diferencias entre distintas razas. Son indicadores de la diversidad dentro de razas, por ejemplo, la heterocigosis, el número medio de alelos o el estadístico F_{IS} . Es, sin embargo, para medir diferencias genéticas entre razas para lo que más se han empleado, con los microsatélites como marcador estrella en este tipo de estudios. Mediante las distancias genéticas se puede determinar qué razas presentan más particularidades genéticas que las hagan alejarse del resto. Pueden ser útiles también para deducir la historia filogenética de la especie, y si bien su uso está extendido por toda la bibliografía especializada, también su idoneidad está sujeta a debate. Por un lado, existen multitud de distancias, y en muchas ocasiones no está claro cuál es la más apropiada para cada caso concreto. El procedimiento de construcción de filogenias y árboles jerárquicos tampoco es único, y distintos métodos con iguales distancias proporcionan topologías que en ocasiones poco tienen que ver entre sí. Por otra parte, un estudio de diversidad basado en distancias genéticas ignora completamente cualquier tipo de variabilidad dentro de poblaciones, puesto que sólo tiene en cuenta las diferencias entre ellas consideradas como un conjunto unitario. Y lo que es más, las conclusiones obtenidas en cuanto a qué razas aportan más diversidad suelen resultar opuestas a las obtenidas cuando se valora qué razas poseen más diversidad en términos de heterocigosis, tamaño efectivo u otras medidas dentro de cada raza.

2.1.- Heterocigosis y estadísticos F

La heterocigosis mide la proporción de individuos heterocigotos existentes para un determinado marcador, o, de forma extendida, su promedio para un conjunto de marcadores. Recordemos que un individuo se considera heterocigoto para un locus cuando éste presenta dos variantes alélicas distintas en el genotipo del individuo, es decir, si para un locus A se pueden dar los alelos $\{A_i\}_{i=1,\dots,k}$, un individuo A_iA_j será heterocigoto si $i \neq j$, y homocigoto en caso contrario. La heterocigosis es una buena medida de la diversidad contenida en una población, ya que está íntimamente relacionada con la consanguinidad y la viabilidad de la población.

La heterocigosis se puede medir de dos formas distintas. Una de ellas consiste en calcular directamente la proporción de individuos heterocigotos contando uno a uno los genotipos heterocigotos, o acumulando las frecuencias génicas de éstos. Es lo que se conoce como *heterocigosis observada*. Así,

para un locus l , $H_{o_l} = \sum_i \sum_{j \neq i} \frac{n_{l_{ij}}}{n}$, siendo $n_{l_{ij}}$ el número observado de

heterocigotos A_iA_j en una población de tamaño n . La proporción poblacional se suele estimar por la proporción muestral, y detalles de las varianzas del estimador dentro y entre poblaciones se pueden encontrar en Weir (1996). Si se consideran m loci, la heterocigosis observada media es simplemente el

promedio de las heterocigosis de cada locus: $H_o = (1/m) \sum_{l=1}^m H_{o_l}$.

La *heterocigosis esperada*, o *diversidad génica*, es la heterocigosis que se esperaría en condiciones de equilibrio Hardy-Weinberg. En poblaciones de censo grande (teóricamente infinito), bajo condiciones de apareamiento aleatorio y ausencia de selección, mutación y migración, los dos alelos dentro

de un locus se comportan de forma estadísticamente independiente y existe lo que se conoce como *equilibrio Hardy-Weinberg*. En esas condiciones, las frecuencias génicas se pueden calcular como el producto de las frecuencias alélicas, y para un locus bialélico, $p_{11} = p_1^2$, $p_{22} = p_2^2$, $p_{12} = 2p_1p_2$, con p_{ij} la frecuencia del genotipo ij y p_i la del alelo i . Por tanto, para ese mismo locus, la heterocigosis esperada se calcularía como $H_e = 2p_1p_2 = 1 - p_1^2 - p_2^2$. En general resulta más cómodo calcularla utilizando las frecuencias de homocigotos, y así, para un locus l con k_l alelos, se tendría que $H_{e_l} = 1 - \sum_{i=1}^{k_l} p_{l_i}^2$, y promediando para

$$m \text{ loci, } H_e = 1 - (1/m) \sum_{l=1}^m \sum_{i=1}^{k_l} p_{l_i}^2.$$

En condiciones de equilibrio $H-W$, $H_o = H_e$, y se pueden construir estadísticos basados en comparar estas dos cantidades para contrastar esa hipótesis. Las diferencias entre la heterocigosis observada y la esperada pueden ser consecuencia de fenómenos poblacionales como la subdivisión y los cuellos de botella.

Una población subdividida muestra generalmente niveles observados de heterocigosis menores a los esperados, y esa reducción en heterocigosis observada puede ser utilizada para cuantificar el grado de diferenciación entre las subpoblaciones. Esta cuantificación se realiza por medio de unos coeficientes conocidos como estadísticos F (Wright, 1951; Cockerham, 1969). Según la denominación de Wright, los estadísticos F se denotan F_{IS} , F_{ST} y F_{IT} , mientras que bajo la notación de Cockerham serían, respectivamente, f , θ y F . Ambas notaciones son correctas y ampliamente utilizadas en la bibliografía. Para definirlos, se consideran en primer lugar las siguientes cantidades:

H_I : heterocigosis *observada* media por individuos dentro de poblaciones

H_S : heterocigosis *esperada* media dentro de poblaciones

H_T : heterocigosis *esperada* media en la población conjunta (calculada con frecuencias alélicas promediadas entre todas las subpoblaciones)

Así, se definen entonces los estadísticos F como

F_{IS} o coeficiente de endogamia: $F_{IS} = f = \frac{H_s - H_I}{H_s}$. Mide la reducción

media de heterocigosis de un individuo debido a apareamientos no aleatorios dentro de una subpoblación, es decir, da una medida del nivel de endogamia dentro de las subpoblaciones. Varía entre -1 (todos los individuos son heterocigotos) a $+1$ (ausencia de heterocigotos observados). Un valor de 0 indicaría igualdad de heterocigotos observados y esperados, y por tanto, una situación de equilibrio.

F_{ST} o índice de fijación: $F_{ST} = \theta = \frac{H_T - H_s}{H_s}$. Es el más utilizado de los

tres, y mide la reducción media de heterocigosis de una subpoblación relativa a la población total debida a la deriva genética entre subpoblaciones. Es, por tanto, una medida del grado de diferenciación genética existente entre subpoblaciones. Representa el porcentaje de la variación genética que es atribuible a las diferencias *entre* poblaciones, y su complementario, por tanto, la proporción atribuible a diferencias *dentro* de poblaciones. Varía entre 0 (no existe diferenciación) a 1 (diferenciación completa, i.e., subpoblaciones fijadas para distintos alelos). Wright (1978) sugirió unas pautas generales para la interpretación del F_{ST} , de acuerdo con las cuales:

$0 \leq F_{ST} < 0.05$ indicaría poca diferenciación genética.

$0.05 \leq F_{ST} < 0.15$ indicaría una diferenciación moderada.

$0.15 \leq F_{ST} < 0.25$ indicaría una gran diferenciación.

$0.25 \leq F_{ST}$ indicaría una diferenciación muy grande.

F_{IT} o índice de fijación conjunto: $F_{IT} = F = \frac{H_T - H_I}{H_I}$. Representa la

reducción media en heterocigosis de un individuo relativa al total poblacional.

Estos tres parámetros están relacionados entre sí a través de una ecuación que permite calcular un tercero conocidos dos de ellos:

$$(1-F_{IT}) = (1-F_{IS})(1-F_{ST}).$$

Esta forma de presentar estos parámetros es sólo una de las varias que hay de conceptualizarlos, en especial el F_{ST} . Distintos autores han formulado el F_{ST} en términos de:

- La varianza de la frecuencia alélica (V_q), la derivación original de Wright (1951). Según ella, $F_{ST} = V_q / pq$.
- La pérdida de heterocigosis a lo largo del tiempo en subpoblaciones sin migración: $F_{ST} = 1 - (1 - 1/2N)^t$.
- Las probabilidades de identidad por descendencia: $F_{ST} = (f_0 - f) / (1 - f)$, donde f_0 es la probabilidad de IBD (Identity By Descent, identidad por descendencia) de dos alelos de la misma subpoblación, y f es la probabilidad de IBD de dos alelos de la población conjunta.
- Los tiempos medios de coalescencia de las subpoblaciones (Slatkin, 1991): $F_{ST} = (t - t_0) / t$, donde t es el tiempo medio de coalescencia para la población conjunta y t_0 el tiempo medio de coalescencia dentro de una subpoblación.
- Un enfoque de tipo ANOVA para descomponer la varianza genética total entre componentes de varianza dentro de y entre subpoblaciones (Weir y Cockerham, 1984).

2.2.- Distancias genéticas

La forma más inmediata y extendida de medir la diversidad genética existente entre dos o más unidades es a través de las distancias genéticas. Cuando una unidad se separa en dos o más, éstas evolucionan a partir de entonces de forma más o menos independiente (tanto más independientemente cuanto menor transferencia genética haya entre ambas en el transcurso de las generaciones). Las fuerzas que actúan sobre cada población (mutación, selección, deriva y migración) irán conformando la composición genética de cada una de ellas, y a medida que aumente el tiempo en generaciones desde la escisión, más se diferenciarán unas de otras.

Lo que se busca es, pues, medir el grado de diferenciación existente entre las distintas poblaciones, razas o especies que se tienen en el momento actual. Como además generalmente no se conoce el proceso de escisiones sucesivas, o filogenia, que ha dado lugar a la composición que se tiene en ese momento, se puede además hacer inferencia filogenética, tanto para conocer la topología en forma de árbol que forman estas escisiones, como para estimar el número de generaciones transcurridas hasta encontrar el origen común de dos ramas.

Para ello se puede definir una distancia genética entre pares de unidades, de forma que cuanto mayor sea la distancia entre ellas, más separadas estarán en términos evolutivos, más distintas serán genética y (presumiblemente) fenotípicamente y mayor será la diversidad entre razas de la que se dispondrá. No existe, sin embargo, una distancia genética única y universal. El tipo de información de la que se dispone, marcadores moleculares de ADN, permite definir diferentes distancias, que presentarán a su vez diferentes propiedades, y cada problema concreto precisará de la elección de

la distancia concreta que mejor se ajuste a sus características: tipo de marcadores, tipo de información sobre ellos, modelo mutacional subyacente, tiempos evolutivos largos (especies) o cortos (razas), o el tipo de inferencia a realizar (topológica o tiempos de divergencia).

La mayoría de las medidas de distancia genética clásicas están basadas en el modelo con infinitos alelos (IAM –Infinite Allele Model-) de Kimura y Crow (1964), o bien en consideraciones geométricas multidimensionales sin referencia a un modelo evolutivo concreto (Goldstein y Pollock, 1997). Si bien el IAM parece ajustar razonablemente el comportamiento de marcadores genéticos tradicionales como polimorfismos de grupos sanguíneos y proteínas (Nei, 1987), la irrupción de los microsatélites como marcadores elegidos en la mayoría de estudios sobre diversidad genética (el 90% de 87 proyectos de investigación encuestados en 2003 relacionados con diversidad genética y cálculo de distancias los utilizaron –Baumung et al., 2004-) ha forzado la aparición de otra serie de distancias que explotan más adecuadamente el modelo evolutivo que mejor se ajustaría a este tipo de marcadores, el SMM (Stepwise Mutation Model, Ohta y Kimura, 1973) y sus generalizaciones.

Los microsatélites son unos segmentos cortos de ADN en los que una cadena específica de nucleótidos de una a seis bases se repite secuencialmente hasta un máximo de 60 (Primmer et al. 1996) o incluso 100 veces (Tautz, 1993), si bien en poblaciones de mamíferos la mayoría de microsatélites presentan un número de alelos observados que oscila entre menos de diez y veinte (Laval et al., 2002).

Estos marcadores están ampliamente distribuidos a lo largo de todo el genoma, presentan una tasa de mutación inusualmente elevada, con valores máximos del orden de 10^{-3} a 10^{-5} en distintas especies (Levinson y Gutman, 1987; Jeffreys et al., 1988; Kelly et al., 1991; Henderson y Petes, 1992; Xu et al., 2000a; Zhivotovsky et al., 2003), además un alto nivel de variabilidad, con niveles de heterocigosis medias dentro de razas generalmente superiores a 0.5, con valores extremos por encima de 0.8 (Laval et al., 2002). Esto hace que sean especialmente adecuados para la reconstrucción de filogenias y

evaluación de distancias en taxones con un historial evolutivo relativamente corto, como los son la mayoría de las razas de animales domésticos dentro de una misma especie, resultando menos informativos para taxones más distantes, como cuando se estudian las relaciones entre distintas especies (Goldstein et al., 1995a; Takezaky y Nei, 1996; Laval et al., 2002).

Pese a que el SMM explica razonablemente bien el modelo mutacional de los microsatélites, lo cierto es que no está claro que el comportamiento de estos marcadores se ajuste estrictamente a él (Weber y Wong, 1993; Di Rienzo et al., 1994), con lo que las conclusiones obtenidas para distancias que se deriven de la asunción del modelo son sólo aproximadamente válidas. Sin embargo, a falta de un modelo mutacional más realista, el SMM es la mejor aproximación disponible.

El IAM asume que cada vez que se produce una mutación, el nuevo alelo será distinto de todos los existentes hasta ese momento. El SMM, sin embargo, supone que un elemento mutacional derivará en un estado alélico anterior o posterior al que produjo la mutación, es decir, que un alelo con 7 repeticiones mutará a uno con 6 o a uno con 8. Originalmente se asumió que la mutación únicamente saltaba un solo paso, y la mayoría de estudios sobre la adecuación de las distancias genéticas se realizan sobre esa asunción, pero la realidad es más compleja.

Existen varios factores que se deben tener en cuenta. Por un lado, la tasa de mutación varía entre distintos loci y alelos, estando relacionada con el número de repeticiones (Ellegren, 2004). En segundo lugar, parece claro que las mutaciones de un solo paso no son las únicas que se producen en los microsatélites, encontrándose habitualmente porcentajes no superiores al 15% de mutaciones multipaso (Weber y Wong, 1993; Goldstein y Pollock, 1997; Brinkmann et al., 1998; Ellegren, 2000; Xu et al., 2000b), aunque se han llegado a documentar porcentajes de hasta el 63% (Huang et al., 2002). La contradicción más importante con el SMM es que de seguir fielmente este modelo, el tamaño de los alelos crecería indefinidamente, cuando en la realidad existen cotas superiores para el número de repeticiones que puede alcanzar un

microsatélite (Goldstein y Pollock, 1997; Ellegren, 2004), y aunque existen generalizaciones del SMM para incluir mutaciones multipaso y limitaciones en el número de alelos (Nauta y Weissing, 1996; Nielsen y Palsboll, 1999; Stefanini y Feldman, 2000; Renwick et al., 2001; Dieringer y Schlotterer, 2003), la mayoría de los estudios sobre distancias genéticas únicamente consideran el SMM unipaso. Tampoco se suele tener en cuenta la asimetría positiva en la distribución de las mutaciones existente en los microsatélites, según la cual éstos tienden a mutar preferentemente a alelos de mayor tamaño que a alelos menores, y cuya intensidad podría depender también del tamaño de los propios alelos (Goldstein y Pollock, 1997).

Dado que el cálculo de una distancia entre dos poblaciones se realizará a partir de la información suministrada por una muestra de individuos de ambas, el resultado será una estimación de la distancia real existente. Por tanto, para la evaluación y comparación de distintas distancias genéticas se utilizan los valores esperados y varianzas de éstas, obtenidas teóricamente, bajo la asunción de un determinado modelo evolutivo, o empíricamente, a través de procedimientos de simulación o remuestreo.

En general, las distancias genéticas se construyen intentando que su esperanza sea lo más lineal posible con respecto al tiempo evolutivo transcurrido y que sean lo más precisas posible, en términos de varianzas muestrales reducidas, de forma que reconstruyan lo más fielmente posible las topologías subyacentes a los conjuntos de poblaciones estudiados, y sean capaces además de estimar correctamente los tiempos de divergencia, todo ello basado en un determinado modelo mutacional. Si una distancia se utiliza para estimar tiempos evolutivos, es esencial que su esperanza se incremente de forma lineal con el tiempo, y es beneficioso que su coeficiente de varianza sea reducido, mientras que si el propósito es la reconstrucción de relaciones filogenéticas, la combinación de linearidad y varianza determinará su rendimiento (Goldstein y Pollock, 1997). De hecho se ha sugerido que en estudios que persigan ambos propósitos, deberían usarse distintas distancias para cada uno de ellos (Takesaki y Nei, 1996). Para la inferencia filogenética en particular, la linearidad de las distancias, y por tanto su fiabilidad para

estimar tiempos de divergencia, debería ser un asunto menor con respecto a la buena determinación de la topología (Laval et al., 2001).

Las distancias genéticas más habituales suelen estar basadas en dos tipos de información: frecuencias y tamaños alélicos, desarrollándose estas últimas como respuesta a la utilización cada vez más habitual de los microsatélites. Se detallarán a continuación las distancias más estudiadas y utilizadas, así como sus características más destacables y las circunstancias en que algunas de ellas son preferibles.

En adelante, se denotarán como X e Y a las dos poblaciones cuya distancia queremos conocer, x_{ij} a la frecuencia en la población X del alelo j del locus i , y a y_{ij} a la correspondiente en la población Y .

Distancia de Sanghvi (Sanghvi, 1953)

$$X^2 = 2(1/r) \sum_{i=1}^r \sum_{j=1}^{m_i} (x_{ij} - y_{ij})^2 / (x_{ij} + y_{ij})$$

Distancia de Rogers

$$D_R = (1/r) \sum_{i=1}^r \left[\sum_{j=1}^{m_i} (x_{ij} - y_{ij})^2 / 2 \right]^{1/2}$$

Distancia estándar de Nei (Nei, 1972)

$$D_s = -\ln(J_{xy} / \sqrt{J_x J_y}),$$

donde $J_x = (1/r) \sum_{i=1}^r \sum_{j=1}^{m_i} x_{ij}^2$, $J_y = (1/r) \sum_{i=1}^r \sum_{j=1}^{m_i} y_{ij}^2$, y $J_{xy} = (1/r) \sum_{i=1}^r \sum_{j=1}^{m_i} x_{ij} y_{ij}$.

Bajo el IAM, D_s aumenta linealmente con el tiempo si el equilibrio mutación-deriva se mantiene durante el proceso evolutivo, esto es, $E(D_s) = 2\nu t$,

donde v es la tasa de mutación por locus por generación y t es el tiempo medido en generaciones desde la divergencia de las dos poblaciones.

Distancia mínima de Nei (Nei, 1973a)

$$D_m = (J_x + J_y)/2 - J_{xy}$$

La esperanza de D_m es conocida e igual a $E(D_m) = J(1 - e^{-2vt})$, donde J es la homocigosis esperada de las dos poblaciones. Obsérvese que se puede llegar de D_m a D_S mediante transformaciones logarítmicas, y viceversa mediante transformaciones exponenciales.

G_{ST} (Nei, 1973b)

$$G_{ST} = \frac{H_T - \bar{H}}{H_T},$$

expresada en función de la heterocigosis total de la población,

$$H_T = 1 - (1/r) \sum_{i=1}^r \sum_{j=1}^{m_i} \bar{p}_{ij}^2 \quad (\text{con } \bar{p}_{ij} = (x_{ij} - y_{ij})/2), \text{ y } \bar{H} \text{ la media de las heterocigosis}$$

esperadas en cada población, $\bar{H} = 1/2(H_x + H_y)$ (con $H_x = 1 - J_x$ y $H_y = 1 - J_y$).

Esta distancia es equivalente al F_{ST} , y se puede reescribir como

$$G_{ST} = \frac{1}{2} \frac{D_m}{1 - (1/r) \sum_{i=1}^r \sum_{j=1}^{m_i} \bar{p}_{ij}^2}, \text{ conocida como distancia de Morton (Morton et al.,}$$

1971). Otras variaciones de la distancia mínima de Nei son la distancia de Latter y la distancia de Reynolds.

La distancia de Reynolds (Reynolds et al., 1983) surge como estimación de otra manera natural de medir distancias entre poblaciones, el coeficiente de parentesco de Malecot (1948), θ , definido como el promedio entre pares de individuos, uno por cada población, de la probabilidad de que dos alelos extraídos al azar, uno de cada individuo, sean idénticos por descendencia. Para

evolución a corto plazo, i.e., t/N pequeño, se tiene que θ sigue una relación aproximadamente lineal con el tiempo, $\theta \approx t/2N$, y aún una mejor aproximación se obtiene con $D = -\ln(1-\theta) \cong t/2N$. Estimando θ mediante componentes de varianza de las frecuencias alélicas dentro y entre poblaciones como la proporción de la varianza total (dentro + entre) atribuible a las diferencias entre poblaciones, se tiene para un locus l que

$$\hat{\theta}_l = \frac{a_l}{a_l + b_l}, \text{ donde } a_l \text{ y } b_l \text{ son los componentes entre y dentro,}$$

respectivamente, con

$$a_l = \frac{1}{2} \sum_{j=1}^{m_l} (x_{lj} - y_{lj})^2 - \frac{(n_1 + n_2)(n_1(1 - J_{x(l)}) + n_2(1 - J_{y(l)}))}{4n_1n_2(n_1 + n_2 - 1)}$$

$$a_l + b_l = \frac{1}{2} \sum_{j=1}^{m_l} (x_{lj} - y_{lj})^2 - \frac{(4n_1n_2 - n_1 - n_2)(n_1(1 - J_{x(l)}) + n_2(1 - J_{y(l)}))}{4n_1n_2(n_1 + n_2 - 1)},$$

siendo n_1 y n_2 los tamaños muestrales en cada una de las poblaciones. Cuando ambos son iguales, una asunción habitual, entonces

$$a_l = \frac{1}{2} \sum_{j=1}^{m_l} (x_{lj} - y_{lj})^2 - \frac{1}{2(2n-1)}(2 - J_{x(l)} - J_{y(l)}), \text{ y } a_l + b_l = 1 - \sum_{j=1}^{m_l} x_{lj}y_{lj}, \text{ con } n$$

el tamaño muestral igual para las dos poblaciones, y para un n suficientemente grande se tiene la fórmula habitual de la distancia de Reynolds:

$$\theta = \frac{(J_x + J_y)/2 - J_{xy}}{1 - J_{xy}}$$

Utilizar la distancia de Reynolds en estos términos es equivalente a usar una distancia que dé una medida de la diversidad dentro de razas corregida por la diversidad entre razas (Laval et al., 2002).

Cuando se tienen en cuenta varios loci existen dos alternativas: tomar medias ponderadas y sin ponderar. Así, sin ponderar se tiene

$$\theta_U = (1/r) \sum_{i=1}^r \theta_i, \text{ frente a una media ponderada}$$

$$\theta_W = \frac{\sum_{i=1}^r a_i}{\sum_{i=1}^r (a_i + b_i)}.$$

Latter (1972b, 1973) propuso su distancia Φ^* , que es idéntica a la distancia de Reynolds para un tamaño muestral suficientemente grande. Obsérvese que esa fórmula también puede ponerse en función de la distancia mínima de Nei como $\Phi^* = \frac{2D_m}{J_x + J_y}$.

Otra variación de la distancia mínima fue también propuesta por Latter (1972a) como

$$\gamma_L = \frac{H_B - H}{1 - H},$$

donde para un locus i , $H_B = 1 - \sum_{j=1}^{m_i} x_{ij} y_{ij}$ y H la frecuencia de heterocigotos

media de las dos poblaciones, de forma que $H_B - H = 1/2 \sum_{j=1}^{m_i} (x_{ij} - y_{ij})^2$.

Siguiendo la idea de Reynolds et al. (1983), se podría tomar $-\ln(1-\Phi^*)$ como medida de distancia, pero se ha demostrado que la eficiencia en cuanto a la obtención de topologías correctas es casi la misma o ligeramente inferior que Φ^* , que a su vez es esencialmente idéntica a θ (Takezaki y Nei, 1996).

Distancia de cuerda (Cavalli-Sforza y Edwards 1967)

$$D_C = (2/\pi r) \sum_{i=1}^r \left\{ 2 \left[1 - \sum_{j=1}^{m_i} \sqrt{x_{ij} y_{ij}} \right] \right\}^{\frac{1}{2}}$$

Si se representan las dos poblaciones en la superficie de una hiperesfera multidimensional usando las frecuencias alélicas en el locus i , D_C para ese locus es la longitud de la cuerda entre las dos poblaciones, donde el ángulo θ_i entre ellas viene dado por $\cos(\theta_i) = \sum_{j=1}^{m_i} \sqrt{x_{ij} y_{ij}}$.

En consideraciones geométricas semejantes está basada la distancia D_A de Nei et al. (1983):

$$D_A = (1/r) \sum_{i=1}^r \left[1 - \sum_{j=1}^{m_i} \sqrt{x_{ij} y_{ij}} \right]$$

Finalmente, como distancias diseñadas específicamente para microsatélites, se tienen las siguientes:

Goldstein et al. (1995a)

$$(\delta\mu)^2 = \frac{\sum_{j=1}^r (\mu_{x_j} - \mu_{y_j})^2}{r},$$

donde μ_{x_j} y μ_{y_j} son las medias de las longitudes alélicas del marcador j en las poblaciones x e y , respectivamente.

ASD (Average Square Distance, Goldstein et al., 1995b, Slatkin, 1995)

$$ASD = \frac{\sum_{k=1}^r \sum_{i,j} (i-j)^2 x_{ik} y_{jk}}{r}$$

Ambas distancias, $(\delta\mu)^2$ y ASD tienen valores esperados conocidos, y ambos son lineales con respecto al tiempo bajo el SMM, siendo $E[(\delta\mu)^2] = 2vt$ y $E(ASD) = 2vt + 2(2N - 1)v$, siendo N el tamaño efectivo. Sin embargo, ASD tiene una varianza mayor que la de $(\delta\mu)^2$, además de ser sensible a efectos de cuello de botella por su dependencia en valor esperado del tamaño efectivo, con lo que a efectos de reconstrucción de topologías es menos recomendable.

Shriver et al. (1995)

$$D_{SW} = W_{xy} - (W_x + W_y)/2,$$

$$\text{donde } W_x = \sum_{k=1}^r \sum_{i \neq j} |i - j| x_{ik} x_{jk} / r, \quad W_y = \sum_{k=1}^r \sum_{i \neq j} |i - j| y_{ik} y_{jk} / r, \text{ y}$$

$$W_{xy} = \sum_{k=1}^r \sum_{i \neq j} |i - j| x_{ik} y_{jk} / r$$

Diversos estudios comparativos entre las distintas distancias (Goldstein et al., 1995a; Takezaki y Nei, 1996; Nagamine e Higuchi, 2001; Laval et al., 2002) confirman empíricamente los resultados teóricos sobre la linearidad de las distintas distancias, además de ofrecer conclusiones de tipo práctico sobre la utilización de unas u otras en determinadas circunstancias.

Takezaki y Nei (1996) destacan que D_A y D_C son las más eficientes en términos generales bajo distintas condiciones a efectos de reconstrucción de topologías, pese a estar basadas en distancias geométricas sobre hiperesferas y ser independientes de los modelos mutacionales. D_A ofrece además una relación lineal con el tiempo evolutivo mejor que la de D_C , de forma que D_A podría ser suficiente para varios propósitos, salvo que el tiempo evolutivo sea grande (en cuyo caso disminuye la linearidad). Distancias como D_S y $(\delta\mu)^2$ incrementan sus valores esperados de forma lineal con el tiempo bajo el IAM y el SMM, respectivamente, como era de esperar, pero presentan un error muestral demasiado alto que hace que su eficiencia para estudiar topologías sea baja. Aún así, para tiempos evolutivos grandes, las distancias diseñadas

para microsatélites, como $(\delta\mu)^2$ y D_{SW} funcionan mejor que las medidas tradicionales. Además, su eficiencia comparada a la de otros modelos que tienen en cuenta mutaciones multipaso en situaciones en las que el SMM no se siga estrictamente es similar, con lo que los detalles de los modelos de cambio en las mutaciones de los microsatélites no son relevantes para la reconstrucción filogenética.

El efecto de las mutaciones en la diferenciación de las poblaciones sólo es perceptible a medio plazo, especialmente si las poblaciones son grandes. En una situación típica en la que se plantean programas de conservación de la diversidad genética, se suelen estudiar razas más o menos locales con una historia relativamente reciente, y que generalmente presentan problemas de endogamia debido al aislamiento reproductivo y a tamaños efectivos reducidos. En situaciones así, y más concretamente, en poblaciones con tamaños efectivos inferiores a 50 y con tasas de mutación del orden de 10^{-3} , las mutaciones se pueden ignorar en los modelos al menos durante 200 generaciones (Laval et al., 2002).

Con modelos basados exclusivamente en deriva, las distancias computadas con frecuencias alélicas dependen en última instancia del número de generaciones desde la divergencia y de los tamaños efectivos de las dos poblaciones, además de parámetros habitualmente desconocidos como las frecuencias en la población fundadora. Todo ello hace que la reconstrucción filogenética se pueda ver enturbiada. En este tipo de situaciones, Laval et al. (2002), tras analizar el comportamiento de diferentes distancias, recomiendan el uso preferente de las distancias de Latter (γ_L) y Reynolds.

En conclusión, pese a que las matrices de distancias genéticas entre grupos de razas pueden estar altamente correlacionadas (Nagamine e Higuchi –2001- encontraron coeficientes de correlación entre las distancias de Sanghvi, D_M , D_S , D_A , D_C , θ_U y θ_W superiores a 0.90 en todos los casos y por encima de 0.99 en muchos de ellos), éstas pueden presentar topologías muy distintas, dependiendo además del tipo de algoritmo que se utilice en la confección del

árbol, de forma que es importante estudiar cuál se ajusta mejor al estudio que se esté llevando a cabo de forma individualizada. No se debe perder de vista, además, que en estudios en los que intervengan razas domésticas, la posibilidad de intercambio de material genético entre ellas a lo largo de la historia, es, más que una posibilidad, un hecho manifiesto en casi todos los casos. Por tanto, las distancias genéticas, sin perder su utilidad como herramienta para estudiar la diversidad presente en el conjunto de esas razas y determinar cuáles de ellas están genéticamente más alejadas o más próximas entre sí, dejan de ser útiles para hacer inferencias filogenéticas, porque no hay filogenia que inferir, al estar las distintas escisiones de hipotéticos ancestros comunes cruzándose frecuentemente y no mantenerse el aislamiento reproductivo. Así, de los dendogramas obtenidos mediante la representación gráfica de las matrices de distancias no se puede deducir una línea evolutiva real; pese a ello siguen siendo herramientas importantes para entender la variabilidad genética existente.

2.3.- Otras medidas clásicas de diversidad

Se revisarán aquí brevemente dos medidas que, por su uso extendido y su relevancia, merecen ser destacadas, además de que una de ellas, el tamaño efectivo, será necesaria para el cálculo de las probabilidades de extinción en la sección 3.1.

El término *tamaño efectivo* surge del estudio de procesos dispersivos en poblaciones no ideales. Una población ideal es aquella en la que se parte de una población base inicial suficientemente grande con apareamiento al azar, y ésta se subdivide en muchas subpoblaciones o líneas más pequeñas, el conjunto de las cuales constituye la población total. En cada subpoblación las frecuencias génicas se ven sometidas a procesos dispersivos, y para su estudio se asumen las siguientes condiciones:

- 1.- Ausencia de intercambio genético entre las líneas. No hay, por tanto, migración, y el aislamiento reproductivo es total.
- 2.- No existe solapamiento generacional.
- 3.- El número de reproductores es el mismo para todas las líneas en todas las generaciones.
- 4.- El apareamiento dentro de cada línea es aleatorio, incluyendo la autofecundación de forma igualmente aleatoria.
- 5.- Ausencia de selección
- 6.- Ausencia de mutación

En estas condiciones, el incremento o tasa de endogamia, ΔF , dado por $\Delta F = (F_t - F_{t-1}) / (1 - F_{t-1})$, con F_t la endogamia en la generación t , se puede calcular como $\Delta F = 1/2N$ (Falconer y Mackay, 1996), donde N es el tamaño constante de cada línea. ΔF así definido expresa la tasa de dispersión, y F el efecto acumulado de la deriva genética.

La tasa de endogamia es un parámetro importante en la genética de poblaciones y cuantitativa, y por tanto interesa conocerlo en circunstancias más reales. Ahora bien, como las circunstancias reales no suelen asemejarse a las de la población ideal, se define el tamaño efectivo, N_e , para suplir el lugar de N en la población ideal. N_e es el número de individuos que originarían la varianza de muestreo observada en los genes de la población, o tasa de endogamia, si estuvieran en las condiciones de la población ideal. N_e se puede calcular, por tanto, como $N_e = 1/2\Delta F$. Cuando la estructura de apareamientos es conocida, se puede relacionar el tamaño efectivo con el real a través de distintas expresiones, que si bien en general son complicadas, se pueden aproximar sin mayor pérdida de precisión (Falconer y Mackay, 1996).

Así, cuando se excluye la autofecundación, $N_e \approx N + \frac{1}{2}$.

Si, además, se excluye el apareamiento entre hermanos, $N_e \approx N + 2$.

Cuando existe un número distinto de machos que de hembras reproduciéndose, se tiene que $\frac{1}{N_e} \approx \frac{1}{4N_m} + \frac{1}{4N_f}$, y $N_e \approx \frac{4N_m N_f}{N_m + N_f}$, con N_m el número de machos y N_f el de hembras.

Si hay distintos tamaños en sucesivas t generaciones, se puede aproximar $\frac{1}{N_e} \approx \frac{1}{t} \left[\frac{1}{N_1} + \frac{1}{N_2} + \dots + \frac{1}{N_t} \right]$.

Estos son los casos más sencillos, pero se pueden derivar expresiones para aproximar el tamaño efectivo cuando se consideran distintos números de descendientes por familia (o por padre y madre), cuando existe solapamiento generacional, además de poder considerar eventos de mutación y migración. Para una revisión exhaustiva puede consultarse Caballero (1994).

La *riqueza alélica* es un parámetro muy informativo comúnmente utilizado como indicador de la diversidad de una población (Marshall y Brown, 1975; Goudet, 1995; Petit *et al.*, 1998; Gutiérrez *et al.*, 2005; Foulley y Ollivier, 2006), y no se trata más que de el número medio de alelos por población. Sin embargo, para poder comparar poblaciones entre sí en función de este indicador tendrían que tener tamaños muy parecidos, ya que es muy sensible al tamaño poblacional. Para evitar este sesgo se recurre al proceso de *rarificación* (Hurlbert, 1971; El Mousadik y Petit, 1996; Kalinowski, 2005), de forma que, seleccionado un tamaño de muestra determinado, se corrige el número de alelos de cada población para ese mismo tamaño. Así, se puede calcular la riqueza alélica como

$$A[g] = \sum_i \left[1 - \frac{\binom{N - N_i}{g}}{\binom{N}{g}} \right],$$

donde g es el tamaño de muestra especificado, N es el número de copias alélicas muestreadas, y N_i el número de ocurrencias del alelo i dentro de los N muestreados.

2.4.- Coeficientes moleculares de parentesco

Muchos estudios sobre conservación de la diversidad genética han descrito la diversidad de un conjunto de varias poblaciones dentro de una especie basándose en distancias genéticas (Moazami-Goudarzi *et al.*, 1997; Thaon D'arnoldi *et al.*, 1998; Eding y Laval, 1999; Ruane, 1999; Laval *et al.*, 2000; Cañón *et al.*, 2001; Aranguren-Méndez *et al.*, 2002; García *et al.*, 2005). Alternativamente, se ha sugerido a modo de estrategia global minimizar el parentesco medio entre animales dentro de una población seleccionada como objetivo de conservación (Haig *et al.*, 1990; Frankham, 1994; Johnston y Lacy, 1995, Zheng *et al.*, 1997; Toro *et al.*, 1998; Eding y Meuwissen, 2001; Caballero y Toro, 2002).

Tanto en la genética de poblaciones como en la cuantitativa, las relaciones genéticas existentes entre individuos pueden resumirse en el coeficiente de parentesco de Malecot (1948), calculado a partir de pedigrís mediante el método tabular de Emik y Terrill (1949). El coeficiente de parentesco f_{XY} (denominado indistintamente *coefficient of kinship* o *coefficient of coancestry* en inglés) entre los individuos X e Y se define como la probabilidad de que para un locus determinado, un alelo escogido al azar de X sea idéntico por descendencia a un alelo escogido al azar de Y. La identidad por descendencia (IBD, Identity By Descent) implica que ambos alelos son copias de un único alelo proveniente de un ancestro común, a diferencia de la identidad en estado (IBS, Identity By State, o AIS, Alike In State), según la cual las dos variantes alélicas son iguales, pero pueden provenir de distintos orígenes.

Para poder calcular el coeficiente de parentesco entre dos individuos es necesario disponer de información genealógica suficiente, lo que no siempre es posible, más aún cuando se trata de relacionar dos individuos de poblaciones distintas. Cuando se estudian razas de especies domésticas, es relativamente frecuente disponer de pedigrís, y de esta forma poder trazar un origen común de dos individuos, pero encontrar orígenes comunes de animales de distintas razas es generalmente inviable.

La reciente disponibilidad masiva de marcadores moleculares ha provocado que, con el fin de utilizar los parentescos para medir la diversidad existente dentro de y entre poblaciones, y en última instancia establecer criterios de priorización basados en ellos, se propongan y estudien en profundidad estimaciones moleculares del parentesco genealógico.

2.4.1 Similitud y parentesco molecular

Eding y Meuwissen (2001) y Toro *et al.* (2002) propusieron por separado utilizar la misma medida de parentesco molecular, basada en una aplicación de la definición de Malecot al caso de IBS. El coeficiente de parentesco molecular (Toro *et al.*, 2002) o índice de similitud (Eding y Meuwissen, 2001) entre dos individuos se define entonces como la probabilidad de que para un marcador determinado, un alelo extraído al azar de uno de ellos sea igual a un alelo extraído al azar del otro, sin tener que serlo necesariamente por descendencia. Para un locus l y dos individuos X e Y puede formularse como

$$f_{M_{XY,l}} = 1/4(I_{11} + I_{12} + I_{21} + I_{22}),$$

donde I_{ij} es la variable indicadora que toma el valor 1 cuando el alelo i del locus l en el primer individuo es igual que el alelo j de ese mismo locus en el segundo individuo.

$f_{M_{XY}}$, el promedio a lo largo de todos los loci de $f_{M_{XY,l}}$, es un estimador del coeficiente de parentesco f_{XY} , y en el caso de que los alelos fundadores

sean únicos se puede probar que es un estimador insesgado (Eding y Meuwissen, 2001).

Cuando existen alelos repetidos en los individuos fundadores, el valor esperado del parentesco molecular está sesgado por la probabilidad de AIS. Denotando esta probabilidad por s , se tiene (Lynch, 1998) que para un locus l ,

$$E[f_{M_{XY,l}}] = f_{XY,l} + (1 - f_{XY,l})s$$

Asumiendo que existe una única población fundadora, todos los individuos actuales están relacionados en menor o mayor medida con ella. Se asume además que los parentescos en la población fundadora son cero, de lo que se deduce que la probabilidad de que dos alelos sean AIS pero no IBD es igual a la probabilidad de que dos alelos sean iguales en la población fundadora, es decir, $s = \sum_k p_k^2$, donde p_k es la frecuencia del alelo k en la población fundadora. En este caso en el que la población fundadora es única y común para todas las poblaciones descendientes, la probabilidad s es, además y salvo mutaciones, igual para todas las poblaciones.

El valor esperado de las similitudes entre dos poblaciones será constante a partir de la separación, y en ausencia de otra información, s se puede estimar por la menor de las similitudes entre poblaciones, que será igual al parecido dentro de la población inmediatamente anterior al primer proceso de escisión, identificando de esta forma a esta población con la población fundadora (Eding y Meuwissen, 2001).

2.4.2 Estimadores moleculares del parentesco genealógico

De las expresiones anteriores se deduce que se puede estimar el coeficiente de parentesco para dos individuos X e Y para un locus l como (Lynch, 1998)

$$\hat{f}_{XY,l} = \frac{f_{M_{XY,l}} - s}{1 - s}$$

Cuando s es constante para todos los marcadores, el estimador de f_{XY} se obtiene promediando $\hat{f}_{XY,l}$ entre todos ellos. Si, en cambio, s es variable, se pueden utilizar ponderaciones basadas en el inverso de la varianza de los estimadores para obtener (Eding y Meuwissen, 2001)

$$[2.4.1] \quad \hat{f}_{XY} = \frac{\sum_l \hat{f}_{XY,l} \left(\frac{1 - s_l}{s_l + \hat{f}_{XY,l}(1 - 2s_l) - \hat{f}_{XY,l}^2(1 - s_l)} \right)}{\sum_l \left(\frac{1 - s_l}{s_l + \hat{f}_{XY,l}(1 - 2s_l) - \hat{f}_{XY,l}^2(1 - s_l)} \right)}$$

En un estudio comparativo de distintas medidas de parentesco molecular y de estimaciones moleculares del parentesco genealógico, Toro *et al.* (2002) concluyen, reforzando la idea ya postulada en Caballero y Toro (2000), que el coeficiente de parentesco molecular, denotado aquí como $f_{M_{XY}}$, es el coeficiente más apropiado para realizar estudios de diversidad y priorización de razas, al ser el que mejor relación presenta con el coeficiente de parentesco de Malecot basado en el pedigrí real (f_{XY}). Su uso se ve nuevamente sugerido en Caballero y Toro (2002), junto con el del parentesco medio entre poblaciones distintas basado en frecuencias alélicas, $f_{ij,l} = \sum_{k=1}^{a_l} \hat{p}_{k,i} \hat{p}_{k,j}$ para un locus l , donde a_l es el número de alelos del locus y $\hat{p}_{k,i}$ y $\hat{p}_{k,j}$ las estimaciones de las frecuencias alélicas del alelo k -ésimo de l en las poblaciones i y j , respectivamente.

Sin embargo, estos resultados están influidos por el hecho de utilizar en el cálculo de $f_{M_{XY}}$ frecuencias alélicas de las poblaciones actuales como si fueran las de la población fundadora, y tal y como se describe en Toro *et al.* (2002) los resultados para \hat{f}_{XY} mejoran cuando se utilizan las verdaderas frecuencias fundacionales. Eding y Meuwissen (2001), mediante la estimación

de s de la forma mencionada anteriormente, obtienen unos resultados satisfactorios, y recomiendan el uso del parentesco estimado por marcadores (*MEK, Marker Estimated Kinship*). Además, bajo la asunción de una única población fundadora, los valores de s serán los mismos para todas las poblaciones finales, lo que tiene como consecuencia que si bien las estimaciones de los parentescos estarán más sesgadas cuanto peor sea la estimación de s , la bondad de esta estima no influirá en una ordenación de las distintas poblaciones en términos de la minimización del parentesco medio.

2.4.3.- Regresión log-lineal en la estimación de parentescos con información molecular

Si consideramos un grupo de N poblaciones, se puede tratar de estimar simultáneamente todos los parentescos para todos los pares de poblaciones y todas las probabilidades de *IBS* para cada locus, obteniendo así una mayor precisión en las estimas y conseguir en última instancia un menor número de contribuciones nulas o negativas en los procedimientos de priorización de poblaciones (ver apartado 3.2). Para ello, Eding y Meuwissen (2002, 2003) propusieron una serie de modelos lineales a partir de transformaciones log-lineales de los datos. Estos modelos se pueden resolver mediante metodología estándar de modelos lineales y obtener volviendo a transformar los resultados, estimas conjuntas de $f_{XY,l}$ y s_l , los coeficientes de parentesco y probabilidades de *IBS* para cada locus y cada par de poblaciones X e Y .

Volviendo a la fórmula que relaciona el parentesco genealógico con el molecular y la probabilidad *IBS*, introduciendo diferentes probabilidades *IBS* por cada locus, se tiene que para cada par de poblaciones i y j , $i, j = 1, \dots, N$, y cada locus l , $l = 1, \dots, L$,

$$E[f_{M_{ij,l}}] = f_{ij} + (1 - f_{ij})s_l.$$

A partir de esta expresión se pueden construir varios modelos de regresión log-lineal. Restando ambos miembros de la igualdad de 1 y tomando logaritmos, se tiene:

$$\ln(1 - f_{M_{ij,l}}) = \ln(1 - f_{ij}) + \ln(1 - s_l) + \varepsilon_{ij,l},$$

o, equivalentemente,

$$y_{ij,l} = a_{ij} + b_l + \varepsilon_{ij,l}.$$

Considerando las $r = \frac{1}{2}N(N+1)$ posibles combinaciones de poblaciones y los L loci, en notación matricial se tiene

$$y = X_a a + X_b b + e,$$

donde y y e son vectores de tamaño $r \cdot L$, a es un vector de tamaño r , b un vector de tamaño L , y X_a y X_b son las matrices del diseño, de tamaños $r \cdot L \times r$ y $r \cdot L \times L$, respectivamente. Si I_k es la matriz identidad de tamaño k y 1_k el vector de tamaño k con unos en todas sus posiciones, entonces $X_a = 1_L \otimes I_r$ y $X_b = I_L \otimes 1_r$, siendo \otimes el producto de Kronecker de dos matrices, lo que corrige el error en Eding y Meuwissen (2003). En un ejemplo con dos poblaciones y dos loci, tendríamos una matriz de parentescos $\begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix}$, de la que, manteniendo la notación, se tendría el vector $a = (a_{11}, a_{12}, a_{22})^t$, y el modelo quedaría establecido como

$$\begin{pmatrix} y_{11,1} \\ y_{12,1} \\ y_{22,1} \\ y_{11,2} \\ y_{12,2} \\ y_{22,2} \end{pmatrix} = \left[\begin{pmatrix} 1 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right] \begin{pmatrix} a_{11} \\ a_{12} \\ a_{22} \end{pmatrix} + \left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right] \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11,1} \\ \varepsilon_{12,1} \\ \varepsilon_{22,1} \\ \varepsilon_{11,2} \\ \varepsilon_{12,2} \\ \varepsilon_{22,2} \end{pmatrix} =$$

$$\begin{aligned}
&= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} \\ a_{12} \\ a_{22} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11,1} \\ \varepsilon_{12,1} \\ \varepsilon_{22,1} \\ \varepsilon_{11,2} \\ \varepsilon_{12,2} \\ \varepsilon_{22,2} \end{pmatrix} = \\
&= \begin{pmatrix} a_{11} \\ a_{12} \\ a_{22} \\ a_{11} \\ a_{12} \\ a_{22} \end{pmatrix} + \begin{pmatrix} b_1 \\ b_1 \\ b_1 \\ b_2 \\ b_2 \\ b_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11,1} \\ \varepsilon_{12,1} \\ \varepsilon_{22,1} \\ \varepsilon_{11,2} \\ \varepsilon_{12,2} \\ \varepsilon_{22,2} \end{pmatrix}.
\end{aligned}$$

A partir de este modelo inicial, se han propuesto tres tipos de regresiones y sus soluciones correspondientes (Eding y Meuwissen, 2003).

– *Modelo loglineal no ponderado (ULM, Unweighted Log-linear Model)*

Consiste en resolver el modelo tal y como se plantea inicialmente. Así, se tiene que las estimaciones de a y b vienen dadas por

$$\begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} X_a^t X_a & X_a^t X_b \\ X_b^t X_a & X_b^t X_b \end{pmatrix}^{-1} \begin{pmatrix} X_a^t y \\ X_b^t y \end{pmatrix}.$$

Las interdependencias existentes en las matrices de diseño llevan a que la solución se obtenga mediante una inversa generalizada de la matriz de coeficientes, con lo que las soluciones para a y b no son únicas.

Además, la solución de a se puede restringir de forma que el menor de los parentescos estimados se fije a cero (Eding *et al.*, 2002). Concretamente,

$$y_{ij,l} = a_{ij} + b_l + \varepsilon_{ij,l} = (a_{ij} - \alpha) + (b_l + \alpha) + \varepsilon_{ij,l},$$

con $\alpha = \max(a_{ij})$.

– *Modelo loglineal ponderado (WLM, Weighted Log-linear Model)*

De la misma forma que en el modelo más sencillo se calculaba una media de las estimaciones del parentesco a lo largo de todos los loci ponderadas por sus varianzas, también ahora se puede plantear una estrategia similar para reducir los errores de estimación tomando en cuenta las varianzas esperadas de los parentescos moleculares. De esta forma, los marcadores más informativos tendrán una mayor influencia en la solución, lo que lleva a mejores estimaciones de los parentescos medios dentro y entre poblaciones. Así, se busca resolver las ecuaciones del modelo, que ahora serán

$$\begin{pmatrix} X_a'W^{-1}y \\ X_b'W^{-1}y \end{pmatrix} = \begin{pmatrix} X_a'W^{-1}X_a & X_a'W^{-1}X_b \\ X_b'W^{-1}X_a & X_b'W^{-1}X_b \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix},$$

de forma que todas las observaciones están ponderadas por la matriz W^{-1} , matriz diagonal (las observaciones se asumen independientes, por lo que los elementos fuera de la diagonal son cero) que contiene los inversos de las varianzas de cada observación, siendo éstas (Eding y Meuwissen, 2003)

$$\text{var}[\ln(1 - f_{M_{ij,l}})] \approx \frac{\hat{f}_{ij} + (1 - \hat{f}_{ij})\hat{s}_l}{4n_{ij}(1 - \hat{f}_{ij})(1 - \hat{s}_l)},$$

siendo n_{ij} el número de observaciones (pares de individuos) realizadas para las poblaciones i y j .

Al resultar las ponderaciones función de los estimadores cuyo valor se pretende hallar, la ecuación del modelo ha de resolverse de forma iterativa.

– *Modelo loglineal mixto ponderado (WLMM, Weighted Log-linear Mixed Model)*

El modelo se diseña para evitar en mayor medida la aparición de contribuciones negativas en un posterior uso de las estimaciones para obtener la aportación de cada población a un conjunto de parentesco mínimo (ver sección 3.2), regresando los *MEK* sobre sus respectivas medias utilizando métodos de modelos mixtos. Para ello se introducen las medias de los estimadores de los parentescos entre y dentro de poblaciones como términos extra:

$$y = Z\mu + X_a a^* + X_b b + e,$$

donde μ es un vector de dos posiciones que contiene las medias de los estimadores entre y dentro de poblaciones en escala logarítmica, Z es la matriz de diseño para μ , de tamaño $r \times 2$, y a^* contiene desviaciones con respecto a la media, de forma que $a = Z\mu + a^*$. Las ecuaciones del modelo serán, por tanto,

$$\begin{pmatrix} Z^t W^{-1} y \\ X_a^t W^{-1} y \\ X_b^t W^{-1} y \end{pmatrix} = \begin{pmatrix} Z^t W^{-1} Z & Z^t W^{-1} X_a & Z^t W^{-1} X_b \\ X_a^t W^{-1} Z & X_a^t W^{-1} X_a + \lambda I & X_a^t W^{-1} X_b \\ X_b^t W^{-1} Z & X_b^t W^{-1} X_a & X_b^t W^{-1} X_b \end{pmatrix} \begin{pmatrix} \mu \\ a^* \\ b \end{pmatrix},$$

donde el término λI se incluye para regresar las desviaciones con respecto a las medias, a^* , de vuelta a cero, y así a vuelve a incluir su media. Cuando λ es igual a cero, este modelo coincide con el anterior, mientras que si tiende a infinito, la matriz de parentescos resultante será tal que sus elementos diagonales serán iguales a μ_1 y los de fuera de la diagonal serán iguales a μ_2 , lo que conllevaría que la solución al problema de las contribuciones de las poblaciones a una metapoblación con parentesco mínimo fuera aquella en la que todas las poblaciones contribuyen de forma idéntica e igual a $1/N$, siendo N el número de poblaciones.

Al igual que en el *WLM*, las ecuaciones deben resolverse por procedimientos iterativos para poder hallar las ponderaciones y el valor de λ .

También se puede escoger λ para que todas las contribuciones al conjunto de parentesco mínimo sean mayores o iguales que cero.

Finalmente, en cualquiera de estos tres modelos loglineales, una vez obtenidas la estimación de a , se deshace la transformación logarítmica para obtener las r estimaciones de los parentescos mediante $\hat{f}_{ij} = 1 - e^{a_{ij}}$.

Los estudios realizados con poblaciones simuladas parecen indicar que el modelo en el que se estiman las probabilidades de *IBS* de forma separada, dado por la ecuación [2.4.1], proporciona unos valores de parentescos estimados que correlacionan con los parentescos genealógicos más que cualquiera de los modelos de estimación simultánea cuando el número de poblaciones es pequeño o el número de marcadores es alto. El *WLM* ofrece mejores resultados cuando se dispone de pocos marcadores y un número mayor de poblaciones, aunque todos los modelos proporcionaron correlaciones superiores a 0.9 en casi todos los casos. En cualquier caso, la flexibilidad de los modelos lineales para incorporar otro tipo de informaciones aconseja su uso para adaptarse mejor a las desviaciones que las poblaciones reales presentan con respecto a situaciones simuladas más o menos ideales. Además, el *WLM* y el *WLMM* se comportan mejor cuando se trata de buscar contribuciones de las distintas poblaciones a la diversidad conjunta, como se verá en la sección 3.2.

2.5.- Diversidad de Weitzman

Una de las medidas de diversidad más utilizadas recientemente (Thaon d'Arnoldi *et al.*, 1998; Laval *et al.*, 2000; Barker *et al.*, 2001; Cañón *et al.*, 2001; Simianer *et al.*, 2003; Reist-Marti *et al.*, 2005; García *et al.*, 2005; Ollivier *et al.*, 2005; Roosen *et al.*, 2005; Solis *et al.*, 2005; European Cattle Genetic Diversity Consortium, 2006; Reist-Marti *et al.*, 2006; Tapio *et al.*, 2006) es la definida por Weitzman (1992), una medida de diversidad que cumple una serie de propiedades deseables, además de proporcionar interpretaciones gráficas razonables a la evolución de ese conjunto de poblaciones. El algoritmo recursivo en el que se basa su cómputo da pie, además, a la implementación de forma relativamente sencilla de los cálculos en un programa informático.

Weitzman (1993) describió también una forma de ponderar las distintas poblaciones con vistas a la distribución de fondos tras la inclusión de éstas en un programa de gestión de recursos genéticos. Este procedimiento se basa a su vez en una medida de diversidad, si bien no necesariamente en la propuesta por él mismo.

La medida de diversidad propuesta por Weitzman fue muy bien acogida desde la perspectiva de los economistas dado que permite definir, sistematizar, y regular algo cuyo concepto hasta entonces no había sido formalizado en ese ámbito, y fue estudiada en mayor profundidad y ampliada en diversos estudios (Glover *et al.*, 1995; Metrick y Weitzman, 1998; Weitzman, 1998; Weikard, 2002), pero no fue hasta unos años después que fue introducida en el ámbito de los genetistas, a través de Thaon d'Arnoldi *et al.* (1998). Las buenas propiedades y la operatividad de esta medida la hacen ciertamente atractiva. Aunque originalmente fue presentada en un contexto en el que el nivel taxonómico a analizar era el de las especies, el principio es aplicable a

cualquier nivel (individuo, población, línea, raza...), pero presenta dos hándicaps cuando la unidad de análisis está por debajo de la especie. Uno es que el modelo no contempla la información sobre la diversidad dentro de cada taxón, sólo tiene en cuenta la variabilidad entre unidades. Esto puede tener sentido cuando se habla de comparar y priorizar especies, pero es difícil de justificar cuando se trabaja, por ejemplo, con razas. Además asume que tras un proceso de fisión y posterior evolución, dos unidades distintas permanecerán aisladas reproductivamente, sin intercambiar material genético. De nuevo, esto es asumible en un contexto de especies, pero no así cuando se habla de razas o líneas.

Pese a todo, la medida de Weitzman y las conclusiones que de ella se derivan para la conservación de diversidad genética siguen resultando de interés, y como se ha citado anteriormente, siguen constituyendo un elemento de análisis válido y utilizado con frecuencia.

2.5.1.- Definición y propiedades.

Como ya se ha mencionado, aunque la medida de diversidad se aplica en la mayor parte de los casos sobre un conjunto de poblaciones, su definición es independiente de la naturaleza del conjunto sobre el que se aplica. En la práctica, sujetos de aplicación de esta medida pueden ser poblaciones, razas, especies, subespecies o incluso sujetos individuales. En el contexto más general se dice que se aplica sobre las denominadas *Unidades Taxonómicas Operativas* (OTUs).

Sea entonces Q un conjunto de OTUs, con $|Q| = N$, y sea $d: Q \times Q \longrightarrow \mathbb{R}^+$ una medida de distancia o disimilitud entre elementos de Q , esto es, d verifica que

$$d(i,j) \geq 0, d(i,j) = d(j,i) \quad \forall \{i,j\} \in Q$$

$$d(i,i) = 0 \quad \forall i \in Q$$

Obsérvese que no se exige la desigualdad triangular, así que el procedimiento es aplicable a distancias no métricas. Por otra parte, en la mayoría de las ocasiones las distancias que se utilicen serán distancias genéticas basadas en información procedente de marcadores moleculares.

El objetivo es, pues, construir una medida de diversidad $D: \wp(Q) \longrightarrow \mathbb{R}^+$ que resulte apropiada, es decir, que verifique una serie de condiciones deseables, a partir de un conjunto de pares de distancias $\{d(i,j)\}_{i,j=1,\dots,N}$.

Si se define la distancia de un elemento a un subconjunto como

$$\begin{aligned} d: Q \times \wp(Q) &\longrightarrow \mathbb{R}^+ \\ (j, S) &\longmapsto d(j, S) := \min_{i \in S} \{d(i, j)\} \end{aligned}$$

sería deseable que D verificase que

$$[2.5.1] \quad D(S \cup j) - D(S) = d(j, S) \quad \forall \emptyset \subset S \subset Q, \quad \forall j \in Q \setminus S.$$

Desafortunadamente, en general [2.5.1] no se cumple. Sin embargo, hay un caso particular en el que la igualdad sí se tiene, y es cuando la distancia d es *ultramétrica*, esto es,

$$\max\{d(i, j), d(j, k), d(i, k)\} = \max\{d(i, j), d(j, k), d(i, k)\} \quad \forall i, j, k \in Q$$

o, equivalentemente, dados $i, j \in Q$ cualesquiera,

$$d(i, j) \leq \max\{d(i, k), d(j, k)\} \quad \forall k \in Q$$

En este caso se puede probar que existe una función $D(Q)$ que verifica [2.5.1], y que esta función se puede expresar como

$$D(Q) = L(Q) + K,$$

con K una constante y $L(Q)$ es la longitud del árbol taxonómico asociado a Q (Weitzman, 1992).

El problema es que en la realidad las distancias muy raramente van a tener propiedades ultramétricas, por lo tanto hay que buscar una medida de diversidad que cumpla unas condiciones similares, pero menos restrictivas.

Como tal medida se considerará la definida por la siguiente fórmula recursiva:

$$[2.5.2] \quad D(Q) = \max_{i \in Q} \{D(Q \setminus i) + d(i, Q \setminus i)\},$$

que tiene solución única una vez fijados $D(i) = d_0 \quad \forall i \in Q$, para un d_0 arbitrario (Weitzman, 1992).

La elección de la solución de [2.5.2] como medida de diversidad tiene una motivación esencialmente pragmática, puesto que como se verá más adelante verifica una serie de condiciones deseables para una función de estas características. Además, la imposición de dos de estas condiciones conlleva de por sí que la medida deba ser coincidente con [2.5.2].

Una de ellas es la *monotonidad en las poblaciones*: cuando se añade un elemento j al conjunto de poblaciones S , $S \subset Q$, la diversidad de S se incrementa al menos en la cantidad $d(j, S)$, esto es,

$$[2.5.3] \quad D(S \cup j) \geq D(S) + d(j, S) \quad \forall \emptyset \subset S \subset Q, \quad \forall j \in Q \setminus S,$$

que es una condición más débil que [2.5.1].

Esta condición se puede restringir añadiendo la llamada *propiedad del enlace*, que garantiza que para todo $\emptyset \subset S \subset Q$ existe al menos una población $j_S \in Q \setminus S$ que satisface

$$[2.5.4] \quad D(S) = D(S \setminus j_S) + d(j_S, S \setminus j_S),$$

es decir, existe alguna población cuya eliminación reducirá la diversidad en la misma magnitud que la distancia de esa población a la más próxima.

Así las cosas se prueba que una medida de diversidad definida según [2.5.2] debe verificar [2.5.3] y [2.5.4], y, recíprocamente, que una medida que cumpla con [2.5.3] y [2.5.4] debe ser acorde con [2.5.2] (Weitzman, 1992).

Otras propiedades que se verifican son las siguientes (Weitzman, 1992):

– *Propiedad de los idénticos*: añadir una población idéntica a otra existente en un conjunto de poblaciones no proporciona ningún incremento de la diversidad. Es decir, sea $\emptyset \subset S \subset Q$ y sea $k \in Q \setminus S$ tal que existe $j \in S$ tal que $d(j, k) = 0$ y $d(j, i) = d(k, i)$ para todo $i \in S$. Entonces $D(S \cup k) = D(S)$.

– *Continuidad de las distancias*: pequeños cambios en las distancias genéticas dan lugar a pequeños cambios en la diversidad. Sea Q' tal que $|Q'| = |Q|$ y $\varphi: Q \rightarrow Q'$ una aplicación biyectiva. Entonces, $\forall \varepsilon > 0 \exists \delta > 0$ tal que si

$$\sum_{i,j} |d(i, j) - d(\varphi(i), \varphi(j))| < \delta,$$

entonces

$$|D(Q) - D(Q')| < \varepsilon.$$

– *Monotonidad en las distancias*: si se incrementan todos los pares de distancias también se producirá un incremento de la diversidad. Sean Q' tal que $|Q'| = |Q| \geq 2$, y $\varphi: Q \rightarrow Q'$ una aplicación biyectiva tal que

$$d(\varphi(i), \varphi(j)) \geq d(i, j) \quad \forall \{i, j\} \in Q.$$

Entonces se tiene que $D(Q') \geq D(Q)$.

2.5.2.- Formulación alternativa de D y representación gráfica.

La resolución de [2.5.2] es computacionalmente costosa, puesto que llega a requerir n^n operaciones, con $n = |Q|$. Sin embargo, puede llegarse a un algoritmo recursivo alternativo que reduce los cálculos a un orden de 2^n . Para ello se utiliza el *Teorema fundamental de representación* (Weitzman, 1992), del que se deduce que la población que hace máxima la expresión de la ecuación [2.5.2], es decir, el *enlace* de [2.5.4], es una de las dos poblaciones más próximas en el conjunto Q . Para $|Q| \geq 2$ se definen las *funciones de mayor similitud*,

$g: Q \rightarrow Q$ y $h: Q \rightarrow Q$ tales que

$$d(g(Q), h(Q)) = \min_{\substack{i, j \in Q \\ i \neq j}} d(i, j).$$

Se define entonces la función $F: \wp(Q) \longrightarrow \mathbb{R}^+$ tal que $\forall \emptyset \subset S \subseteq Q$ con $|S| = 1$, $F(S) := d_0$, y $\forall \emptyset \subset S \subseteq Q$ con $|S| \geq 2$, $F(S)$ es la solución de la ecuación recursiva

$$[2.5.5] \quad F(S) = d(g(S), h(S)) + \max\{F(S \setminus g(S)), F(S \setminus h(S))\}.$$

El teorema asegura que en este caso, $F(S) = D(S)$ para todo S .

La expresión que da lugar a [2.5.5] es fácilmente implementable en un programa informático y reduce considerablemente el tiempo de computación con respecto a [2.5.2]. Aproximaciones en la forma de calcular [2.5.5] pueden reducir aún más este tiempo (Weitzman, 1992; Thaon d'Arnoldi *et al.*, 1998; García *et al.*, 2005), como se verá más adelante.

Aunque el objetivo inicial de [2.5.5] es proporcionar un valor de diversidad para un conjunto de poblaciones, el algoritmo que conduce a la obtención de este valor es un algoritmo de agrupación, con lo que la información contenida en las distancias genéticas puede representarse gráficamente en forma de un árbol jerárquico, cuya topología es única si se identifican las poblaciones que actúan de enlace. El árbol así construido permite una fácil visualización del efecto a que daría lugar en términos de diversidad la pérdida de una población o un conjunto de ellas.

Debe tenerse en cuenta que este árbol es una representación de la diversidad observada a través de las distancias genéticas y, por lo tanto, matrices de distancias genéticas escasamente correlacionadas darán lugar a topologías muy diferentes. Además, aunque del árbol se puedan obtener conclusiones sobre los clusters que lo forman, salvo raras excepciones, la topología observada no puede ser interpretada como el reflejo del verdadero proceso evolutivo sufrido por las poblaciones objeto de estudio.

2.5.3.- Incorporación de información intrapoblacional.

Como ya se ha comentado, la medida de diversidad de Weitzman tiene el serio inconveniente de no contemplar la variabilidad acumulada dentro de cada unidad (Caballero y Toro, 2002; Eding *et al.*, 2002; Bruford, 2004). Si bien metodológicamente esto no es necesariamente un problema, cuando se aplica a unidades como razas sí que es conveniente tener en cuenta la variabilidad dentro de ellas, en especial cuando la diversidad se utilice para establecer prioridades de conservación, ya que de no hacerlo se estarán priorizando las razas más alejadas, que generalmente son las que más deriva genética han sufrido y, por tanto, menos diversidad retienen. Habitualmente los resultados obtenidos utilizando esta medida de diversidad son opuestos a los que se obtendrían si se consideraran indicadores como la heterocigosis, la endogamia o el tamaño efectivo.

La forma más inmediata de evaluar la diversidad dentro de poblaciones en el contexto de la diversidad de Weitzman es aplicar esta medida en cada raza, contemplando a los individuos como unidades y utilizando distancias genéticas entre individuos (García y Cañón, 2005). Al igual que sucede en cualquier otro nivel (razas, especies, etc.), sólo se tiene en cuenta la diversidad entre unidades, así que una población tendrá mayor diversidad cuanto más alejados estén sus individuos en términos de distancias genéticas. Como el alejamiento es mayor cuanto más homocigotos sean los individuos, el resultado final es que las poblaciones con mayor diversidad de Weitzman serán aquellas con menor heterocigosis. Esto tiene pleno sentido al ser la heterocigosis una medida de diversidad dentro de individuos, y la de Weitzman una medida de diversidad entre individuos. En la Tabla 2.5.1 se muestran resultados de un ejemplo con datos reales de ocho razas de vacuno, y se puede comprobar que las heterocigosis esperadas muestran correlaciones de cerca de -1 con las dos distancias entre individuos estudiadas (ASD y Malecot).

He							
N_e	0,46						
ASD	-0.92	-0.21					
Mal	-0.98	-0.47	0.94				
CP_{Rey}	-0.33	-0.77	-0.05	0.27			
CP_{ReyHe}	-0.16	-0.64	-0.21	0.10	0.97		
CP_{ReyNe}	0.30	0.89	-0.13	-0.32	-0.52	-0.36	

Tabla 2.5.1: Correlaciones entre distintos indicadores de diversidad para ocho razas de vacuno. He: Heterocigosis esperada media, N_e: tamaño efectivo (aproximado), ASD: Diversidad de Weitzman dentro de razas medida con la distancia ASD. Mal: idem, con la distancia de Malecot. CP_{Rey}: contribuciones parciales a la diversidad de Weitzman, calculada ésta con la distancia de Reynolds. CP_{ReyHe}: contribuciones parciales utilizando la distancia de Reynolds ponderada por la heterocigosis esperada. CP_{ReyNe}: contribuciones parciales utilizando la distancia de Reynolds ponderada por el tamaño efectivo.

Dado el consumo computacional que requiere el cálculo de esta diversidad, y que haría prácticamente imposible efectuarlo para una muestra de un tamaño razonable de cualquier población, para obtener buenas estimaciones se sugiere un procedimiento de remuestreo, con un tamaño muestral manejable por el algoritmo y un número de repeticiones suficiente

para tener buenas estimaciones. Además es necesario que el tamaño muestral sea el mismo en todas las poblaciones, ya que el valor de diversidad depende del número de unidades para el que se compute.

Otra forma de incorporar la variabilidad dentro de poblaciones a la diversidad de Weitzman es hacerlo directamente desde su formulación. En García y Cañón (2005) se propuso modificar la fórmula recursiva con la que se calcula la diversidad, introduciendo en cada iteración la variabilidad correspondiente a la población enlace, es decir, que suponiendo que $g(S)$ sea la población que haga máxima la expresión de [2.5.5], es decir, que

$$F(S) = d(g(S), h(S)) + F(S \setminus g(S)),$$

entonces se redefine la diversidad como

$$F'(S) = d(g(S), h(S)) \cdot He(g(S)) + \max\{F'(S \setminus g(S)), F'(S \setminus h(S))\}.$$

A pesar de que parece que el comportamiento de la diversidad con esta nueva definición parece más equilibrado que el de la definición original en cuanto a concordancia con las medidas de diversidad dentro de poblaciones, la elección de $g(S)$ y no $h(S)$ no deja de ser arbitraria, y dota a la medida de una falta de simetría importante.

Para remediar esto, se propone una definición alternativa de diversidad en la que se combinan las variabilidades dentro de las dos poblaciones que tienen distancia mínima. Así, se propone utilizar

$$F'(S) = d(g(S), h(S)) \frac{V(g(S)) + V(h(S))}{2} + \max\{F'(S \setminus g(S)), F'(S \setminus h(S))\},$$

donde $V(i)$ es una medida de la variabilidad dentro de la población i .

Esta formulación alternativa tiene la ventaja de que se puede contemplar como la aplicación de la original en [2.5.5] a una matriz de distancias alternativa en la que

$$d'(i, j) = d(i, j) \frac{V(i) + V(j)}{2}.$$

Obsérvese que d' cumple las tres propiedades enunciadas por Weitzman (1992) para que sea susceptible de ser utilizada en este tipo de análisis, ya que $d'(i, j) \geq 0$, $d'(i, i) = 0$ y $d'(i, j) = d'(j, i)$. Al estar las distancias ponderadas por medidas de diversidad dentro de poblaciones se consigue que las poblaciones que más aporten a la diversidad conjunta sean aquéllas que mejor equilibrio tengan entre separación del resto y variabilidad dentro, aunque como siempre, todo depende en última instancia de las elecciones que se hagan de distancia inicial y medida de diversidad dentro de poblaciones. En la tabla 2.5.1 se puede comprobar para un ejemplo real cómo la correlación de las contribuciones a la diversidad conjunta de cada una de las razas con la heterocigosis y el tamaño efectivo disminuye en su negatividad cuando se utilizan las distancias ponderadas por la heterocigosis media, y cómo incluso es positiva cuando se utiliza como ponderación el tamaño efectivo. Las tablas 2.5.2 y 2.5.3 muestran los valores de heterocigosis, tamaños efectivos, diversidades dentro y contribuciones a la diversidad de cada raza, y las distintas ordenaciones que estos valores producen entre las distintas razas.

Razas	He	N _e	CP _{Rey}	CP _{ReyHe}	CP _{ReyNe}
SAL	0,369 100		8,040	8,040	13,320
AUB	0,389 100		8,490	8,490	16,570
GAS	0,292 100		10,820	10,820	17,930
ALI	0,319 36		16,960	16,960	10,530
ASM	0,295 35		8,740	8,740	6,900
ASV	0,317 89		9,600	9,600	18,610
SAY	0,293 21		18,760	18,760	7,400
TUD	0,349 35		16,860	16,860	10,400

Tabla 2.5.2: Valores de los distintos indicadores de diversidad y contribuciones a la diversidad conjunta de cada una de las razas (ver sección 3.1).

He	N _e	CP _{Rey}	CP _{ReyHe}	CP _{ReyNe}
AUB	SAL	SAY	TUD	ASV
SAL	AUB	ALI	SAY	GAS
TUD	GAS	TUD	ALI	AUB
ALI	ASV	GAS	ASV	SAL
ASV	ALI	ASV	GAS	ALI
ASM	ASM	ASM	AUB	TUD
SAY	TUD	AUB	SAL	SAY
GAS	SAY	SAL	ASM	ASM

Tabla 2.5.3: Orden de importancia de las distintas razas de acuerdo con cada uno de los indicadores.

Como se puede comprobar, parece la ponderación de las distancias genéticas por medidas de diversidad dentro de poblaciones ayuda pues a equilibrar y redistribuir las importancias relativas de cada raza a la diversidad conjunta.

Finalmente, otra forma de tener en cuenta la variabilidad existente en cada población es utilizar información relativa a ella en el cálculo de las probabilidades de extinción para obtener diversidades marginales y elasticidades (Bennewitz y Meuwissen, 2005; García *et al*, 2005; Simianer, 2005), como se verá en el apartado 3.1.

2.5.4.- Aproximación del algoritmo para un número grande de poblaciones

A nivel computacional, el cálculo de la diversidad de Weitzman presenta un problema importante, y es el crecimiento exponencial del algoritmo recursivo de la expresión [2.5.5]. En la Fig. 2.5.1 puede verse un ejemplo desarrollado con cuatro poblaciones del funcionamiento del algoritmo. Partiendo de la matriz de distancias inicial, en cada iteración primero se busca la pareja de unidades con menor distancia entre sí, y a continuación se calcula la diversidad para dos nuevos subconjuntos, los formados por el conjunto inicial excluyendo cada una de las dos unidades, escogiendo el máximo de los dos valores obtenidos. Así, en cada paso el problema se duplica, duplicándose asimismo el tiempo de

computación, que aumenta exponencialmente con el número de unidades iniciales. Dependiendo de las características del equipo informático de que se disponga, esto implica que el cálculo de la diversidad no sea operativo para más de entre 30 y 40 poblaciones.

Inicio

64 Pob. 1

Medidas de diversidad genética

Pob. 2 0.245

Pob. 3 0.153 0.127

Pob. 4 0.271 0.294 0.220

Paso 1: encontrar mínimo
(pobs. 2 y 3)

Pob. 1

Pob. 2 0.245

Pob. 3 0.153 0.127

Pob. 4 0.271 0.294 0.220

$$D = 0.127 + \left\{ \right.$$

Paso 1.1: eliminar
(pob. 2)

Pob. 1

Pob. 2 0.245

Pob. 3 0.153 0.127

Pob. 4 0.271 0.294 0.220

Paso 2 (1.1) : encontrar mínimo
(pobs. 1 y 3)

Pob. 1

Pob. 2 0.245

Pob. 3 0.153 0.127

Pob. 4 0.271 0.294 0.220

$$D = 0.127 + \left\{ \begin{array}{l} 0.153 + \left\{ \end{array} \right.$$

Paso 2.1 (1.1): eliminar
(pob. 3)

Pob. 1

Pob. 2 0.245

Pob. 3 0.153 0.127

Pob. 4 0.271 0.294 0.220

Paso 3 (2.1 (1.1)): encontrar mínimo
(pobs. 1 y 4)

Pob. 1

Pob. 2 0.245

Pob. 3 0.153 0.127

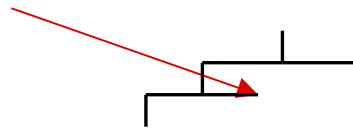
Pob. 4 0.271 0.294 0.220

$$D = 0.127 + \left\{ \begin{array}{l} 0.153 + \left\{ \begin{array}{l} 0.271 \end{array} \right. \end{array} \right.$$

Fig 2.5.1: Representación del algoritmo en un ejemplo simulado con cuatro poblaciones

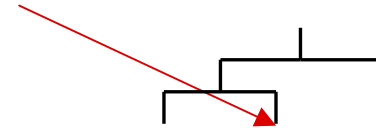
Paso 2.2 (1.1): eliminar (pob. 1)

Pob. 1			
Pob. 2	0.245		
Pob. 3	0.153	0.127	
Pob. 4	0.271	0.294	0.220



Paso 3 (2.2 (1.1)): encontrar mínimo (pobs. 3 y 4)

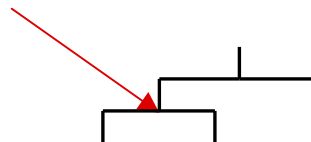
Pob. 1			
Pob. 2	0.245		
Pob. 3	0.153	0.127	
Pob. 4	0.271	0.294	0.220



$$D = 0.127 + \left\{ \begin{array}{l} 0.153 + \left\{ \begin{array}{l} 0.271 \\ 0.220 \end{array} \right\} \end{array} \right.$$

Paso 2 (1.1) : evaluar diversidad

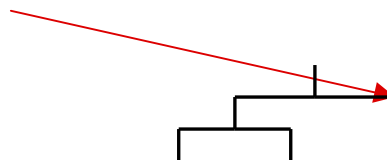
Pob. 1			
Pob. 2	0.245		
Pob. 3	0.153	0.127	
Pob. 4	0.271	0.294	0.220



$$D = 0.127 + \left\{ \begin{array}{l} 0.153 + \left\{ \begin{array}{l} 0.271 \\ 0.220 \end{array} \right\} \end{array} \right. \quad D = 0.127 + \left\{ \begin{array}{l} 0.153 + 0.271 \end{array} \right. \quad D = 0.127 + \left\{ \begin{array}{l} 0.424 \end{array} \right.$$

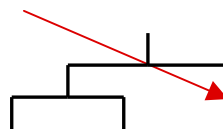
Paso 1.2: eliminar (pob. 3)

Pob. 1			
Pob. 2	0.245		
Pob. 3	0.153	0.127	
Pob. 4	0.271	0.294	0.220



Paso 2 (1.2) : encontrar mínimo (pobs. 1 y 2)

Pob. 1			
Pob. 2	0.245		
Pob. 3	0.153	0.127	
Pob. 4	0.271	0.294	0.220



$$D = 0.127 + \left\{ \begin{array}{l} 0.424 \\ 0.245 + \left\{ \begin{array}{l} \end{array} \right\} \end{array} \right.$$

Fig. 2.5.1 (cont.)

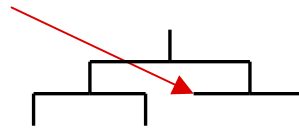
Paso 2.1 (1.2): eliminar (pob. 1)

Pob. 1

Pob. 2 0.245

Pob. 3 0.153 0.127

Pob. 4 0.271 0.294 0.220

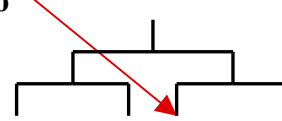
**Paso 3 (2.1 (1.2)): encontrar mínimo (pobs. 2 y 4)**

Pob. 1

Pob. 2 0.245

Pob. 3 0.153 0.127

Pob. 4 0.271 0.294 0.220



$$D = 0.127 + \left\{ \begin{array}{l} 0.424 \\ 0.245 + \left\{ \begin{array}{l} 0.294 \end{array} \right. \end{array} \right.$$

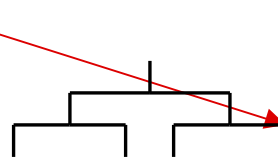
Paso 2.2 (1.2): eliminar (pob. 2)

Pob. 1

Pob. 2 0.245

Pob. 3 0.153 0.127

Pob. 4 0.271 0.294 0.220

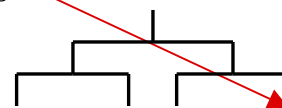
**Paso 3 (2.2 (1.1)): encontrar mínimo (pobs. 1 y 4)**

Pob. 1

Pob. 2 0.245

Pob. 3 0.153 0.127

Pob. 4 0.271 0.294 0.220

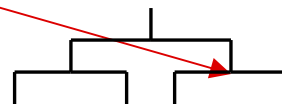
**Paso 2 (1.2) : evaluar diversidad**

Pob. 1

Pob. 2 0.245

Pob. 3 0.153 0.127

Pob. 4 0.271 0.294 0.220



$$D = 0.127 + \left\{ \begin{array}{l} 0.424 \\ 0.245 + \left\{ \begin{array}{l} 0.294 \\ 0.271 \end{array} \right. \end{array} \right.$$

$$D = 0.127 + \left\{ \begin{array}{l} 0.424 \\ 0.245 + 0.294 \end{array} \right.$$

$$D = 0.127 + \left\{ \begin{array}{l} 0.424 \\ 0.539 \end{array} \right.$$

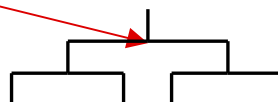
Paso 1: evaluar diversidad

Pob. 1

Pob. 2 0.245

Pob. 3 0.153 0.127

Pob. 4 0.271 0.294 0.220



$$D = 0.127 + \left\{ \begin{array}{l} 0.424 \\ 0.539 \end{array} \right.$$

$$D = 0.127 + 0.539$$

$$D = 0.666$$

Fig 2.5.1 (Cont.)

En la Fig. 2.5.2, procedente de un ejemplo real, se puede ver el ajuste de una regresión log-lineal del tiempo empleado en centésimas de segundo sobre el número de unidades incluidas en el conjunto inicial. Como se puede observar, el ajuste es casi perfecto, con un coeficiente de determinación $R^2=0.9997$. El tiempo t en función del número de poblaciones n vendría dado por $t = 2.7 \cdot 10^{-4} \cdot (2.14)^n$. Para hacerse una idea del tiempo que puede consumir el cálculo de la diversidad, en un ordenador PII a 400 MHz, 29 poblaciones consumieron 2h 50', y 34 poblaciones casi 5 días y 6h, mientras que en un equipo AMDXP a 2.4GHz, el cálculo para 36 poblaciones empleó más de 4 días y 9h. Es claro, pues, que se necesita algún tipo de aproximación que permita manejar un mayor número de poblaciones, especialmente si se quiere calcular la diversidad al nivel de individuos, o para el cálculo de las contribuciones marginales y elasticidades que se estudiarán en el epígrafe 3.1.

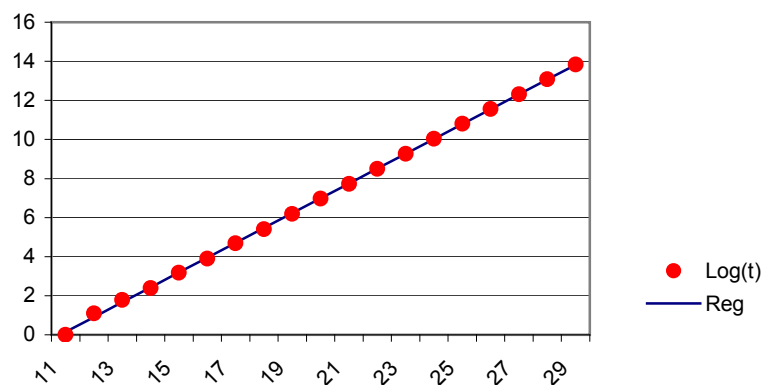


Fig. 2.5.2: Valores del logaritmo del tiempo consumido en cseg. para el cálculo de la diversidad de Weitzman en conjuntos con distintos números de poblaciones, y recta de regresión loglineal ajustada a los datos.

Thaon d'Arnoldi *et al.* (1998) propusieron una aproximación consistente en muestrear un número determinado de caminos de entre los 2^{n-1} que el algoritmo examina en el árbol binario (ver Fig. 2.5.1) y tomar como valor aproximado de la diversidad el máximo de los valores procedentes de los árboles muestreados. Este procedimiento funciona bien con un número de poblaciones relativamente bajo (en su artículo, Thaon d'Arnoldi *et al.* lo testaron hasta con 29 poblaciones, pero a medida que aumenta el número de unidades,

también debe hacerlo el tamaño muestral necesario para tener una buena aproximación. El problema es que mientras que el algoritmo exacto se puede implementar de forma recursiva, el aproximado no, y se llega a un punto a partir de un determinado número de unidades en el que el algoritmo aproximado consume más tiempo que el exacto (García *et al.*, 2005).

En el artículo de García *et al.* se propone otra aproximación bastante más eficiente, al menos en cuanto a tiempo de computación. Como ya se ha visto, el algoritmo se divide en dos en cada paso, repitiéndose recursivamente este patrón. En cada una de esas subdivisiones se analiza una submatriz de distancias consistente en quitar la fila y columna correspondientes a una determinada población de la matriz anterior. Lo que se propone es comprobar la dispersión de cada una de estas submatrices antes de ser analizadas. Si ésta es pequeña, es decir, los valores de la matriz están suficientemente agrupados en torno a su valor medio, se usará este valor, multiplicado por el número de pasos restantes, como valor de la diversidad para esa matriz. En caso contrario, si los valores no están suficientemente concentrados, se procederá con el análisis habitual (exacto) de la matriz. La forma de evaluar esa concentración será determinando un umbral para el coeficiente de variación (desviación estándar dividida por la media) de los valores de la matriz. Si este coeficiente es menor que el umbral elegido, entonces se aproxima ese paso; si es mayor, se prosigue con el algoritmo exacto (que podría truncarse por aproximación en una iteración posterior).

La figura 2.5.3 muestra la diferencia observada entre los valores reales y los aproximados de la diversidad para un número creciente de poblaciones, de 13 a 33. Con un umbral para el coeficiente de variación de 0.25, esta diferencia nunca supera el 3% en valor absoluto. Los tiempos de computación, sin embargo, se reducen drásticamente, pasando de casi dos días y 11h para las 33 poblaciones finales en el algoritmo exacto a 12 minutos y medio en el aproximado, con una diferencia entre los dos valores obtenidos del 2.69% todo ello en un ordenador de tipo PII a 400 MHz. Obviamente, a mayor umbral, mayor velocidad en la computación, puesto que será más fácil que el

coeficiente de variación quede por debajo de él y se lance la aproximación, pero mayor será también la diferencia entre el valor real y el aproximado.

El comportamiento del error en la aproximación, como se muestra en la figura 2.5.3, es bastante irregular, debido a que la aproximación es dependiente de la estructura de los datos (en este caso, de la matriz de distancias). Para un conjunto de datos distinto al utilizado para la figura 2.5.3, 35 poblaciones consumieron casi dos días con el algoritmo exacto, mientras que la aproximación tardó en obtener una solución menos de dos segundos con un umbral de 0.4 y un error respecto al valor real de menos del 2%. Por tanto, pese a su irregularidad y falta de predictibilidad en su comportamiento, parece ofrecer buenos resultados, y, en cualquier caso, parece más útil que la otra aproximación conocida de este algoritmo, la propuesta por Thaon d'Arnoldi *et al.* (1998).

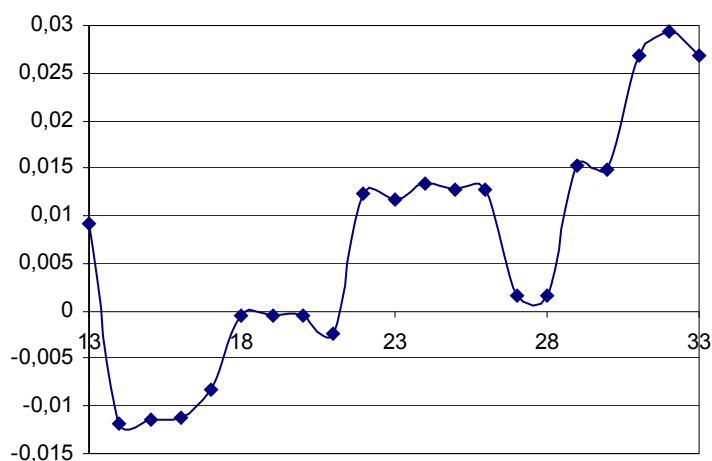


Fig. 2.5.3: Diferencia entre las diversidades reales y las aproximadas para distintos números de poblaciones

3.- Estrategias de asignación de prioridades en esquemas de conservación

Sea *in vivo* o *in vitro*, lo cierto es que cuando se plantea un programa de conservación de la diversidad genética hay que asignar una cantidad determinada de recursos económicos. Tanto la recuperación *in vivo* de una raza como la creación de bancos de germoplasma requieren una inversión. Ahora bien, parece razonable que esa inversión sea proporcional a lo importante que sea esa raza en términos de diversidad dentro del conjunto de todas las razas de la especie que se quiera conservar. El problema es que existe poco consenso dentro de la comunidad científica sobre cómo determinar esa importancia. Existen distintas metodologías que proporcionan resultados a veces contrapuestos, se discute sobre el peso que deben tener los componentes entre razas y dentro de razas, y ni siquiera hay una definición universal de lo que se entiende por diversidad genética.

En ocasiones ni siquiera existe un consenso sobre cómo definir las unidades sobre las que actuar. En términos generales, se puede decir que cuando una población muestra una adaptación diferencial a un determinado medio o diferencias genéticas significativas, estaría justificada su consideración en programas de conservación como una unidad evolutiva. Se ha utilizado el concepto de *unidades evolutivas significativas* (Fraser y Bernatchez, 2001) para referirse a poblaciones aisladas desde el punto de vista reproductivo e histórico, y adaptadas a medios claramente diferentes y, recientemente se ha propuesto la utilización del concepto de *posibilidad de intercambio* o *exchangeability*, (Crandall *et al*, 2000) para clasificar a las poblaciones que pueden constituir una especie. En este sentido las razas se clasificarían en función, por un lado, del grado de intercambio genético y ecológico, y por otro, de si este intercambio se ha producido recientemente o históricamente.

Como se ha visto en los apartados 2.4 y 2.5, en los últimos años se han desarrollado medidas de diversidad más complejas con el objeto de establecer criterios objetivos para la asignación de recursos económicos a las unidades

según la participación de cada una de ellas en la diversidad genética global. La diversidad de Weitzman utiliza distancias genéticas entre pares de poblaciones para obtener una única medida que verifica una serie de propiedades que una buena medida de diversidad debería tener. Pese a su abundante utilización en la bibliografía reciente, presenta el importante handicap de prescindir totalmente en su definición de la información genética *dentro* de poblaciones, ya que se basa exclusivamente en distancias. Se han visto en el apartado 2.5.3 mejoras conceptuales que permiten la inclusión de información intrapoblacional en el propio algoritmo, y a continuación se presentarán formas alternativas de considerarla, mediante probabilidades de extinción genética basadas en tamaños efectivos y frecuencias alélicas de los marcadores (García *et al.*, 2005). Aún así, el procedimiento sigue sujeto a cierta controversia.

La utilización tradicional de coeficientes de parentesco y endogamia medios para determinar el aporte de cada raza a la diversidad global, entendida como concepto opuesto al de endogamia, presentaba problemas prácticos como que la eliminación de una raza provocara de hecho un aumento de la diversidad. Con el objeto de eliminar esta traba y poder obtener una medida de diversidad basada en parentescos que verificara las mismas condiciones que la de Weitzman y utilizara información dentro y entre poblaciones se desarrolló el método del conjunto nuclear (Eding *et al.*, 2002), que contempla la creación hipotética de un conjunto de animales con representación de todas las razas en el que tras su creación, los animales se reproducirían aleatoriamente, de forma que la consanguinidad en ese conjunto fuera mínima, y se definiría la diversidad como el opuesto a esa consanguinidad mínima. Para llegar a ello se cuenta con una matriz de parentescos dentro y entre poblaciones, que en la práctica suelen ser parentescos moleculares, es decir, estimaciones de los parentescos reales obtenidas con información de marcadores moleculares, tal y como se desarrollaron en el apartado 2.4. A partir de ahí, una raza sería más prioritaria cuanto más contribuyera proporcionalmente a ese conjunto nuclear, o cuanto más descendiera la diversidad conjunta al eliminarla del conjunto global. La segunda parte de este apartado estará dedicada al estudio de este procedimiento.

Finalmente, se revisará brevemente una tercera vía propuesta como alternativa para obtener unas contribuciones a la diversidad global que contemplen ambos tipos de información genética (dentro y entre poblaciones), la resultante de obtener contribuciones basadas en información “dentro” por un lado y basadas en información “entre” por otro, y combinarlas en una única contribución, suma de las dos obtenidas por separado, ponderadas por un factor de ponderación que refleja la importancia relativa de la diversidad “entre” con respecto a la “dentro” (Ollivier y Foulley, 2002, 2005).

3.1.- Contribuciones parciales, diversidad esperada, diversidad marginal y elasticidad de la diversidad

Asumiendo que se dispone de una medida de diversidad en un conjunto de poblaciones, interesa conocer como evolucionará esa diversidad en el futuro y qué poblaciones tienen más peso en la conservación de la diversidad genética de cara a una distribución de recursos en programas de gestión de biodiversidad. Para dar solución a esto se proponen los conceptos de *contribuciones parciales*, esbozado en el apartado 2.5.3, *diversidad esperada*, *diversidad marginal* y *elasticidad de la diversidad*.

Sea Q un conjunto de OTUs. Sea $D: \wp(Q) \longrightarrow \mathbb{R}$ una función de diversidad cualquiera. La *Contribución Parcial (CP)* de la unidad i se define como la pérdida de diversidad actual al eliminar la población i del conjunto, i.e.,

$$[3.1.1] \quad CP_i = \frac{D(Q) - D(Q \setminus i)}{D(Q)},$$

siendo $D(Q \setminus i)$ la diversidad del conjunto Q excluyendo la unidad i .

Sean ahora $\{P_i\}_{i=1, \dots, N}$ las probabilidades de extinción con el paso de una generación de cada una de las N poblaciones.

Se define la *Diversidad Esperada Descontada (DE)* al cabo de T generaciones como

$$[3.1.2] \quad DE_T = \sum_{S \in \wp(Q)} P_i(S) D(S),$$

donde $P_t(S)$ es la probabilidad de que el conjunto de poblaciones contenidas en S hayan sobrevivido al cabo de T generaciones y que las contenidas en $Q \setminus S$ se hayan extinguido. Para calcularla se asume que el tiempo de “vida” de cada especie sigue una distribución exponencial de parámetro P_i , de forma que $P_t(S)$ vendría dada por

$$[3.1.3] \quad P_t(S) = \prod_{i|P_i \in S} \exp(-P_i t) \prod_{i|P_i \in Q \setminus S} (1 - \exp(-P_i t)).$$

La fórmula [3.1.2] no coincide exactamente con la propuesta por Weitzman (1993), ya que originalmente se define la diversidad esperada como la suma acumulada de las generaciones 1 a la t , ponderada por un factor de descuento. Este factor se ha ignorado aquí ya que la diversidad esperada será parcialmente diferenciada con respecto a las probabilidades de extinción, y este factor jugaría el papel de una constante sin valor a efectos de comparar las distintas diferenciales. Por otra parte, no hay beneficio alguno en acumular la suma por generaciones, además de que la ecuación [3.1.2] es más coherente con el concepto estadístico de valor esperado (García *et al.*, 2005).

Un problema relativamente importante es que en sentido estricto, la diversidad que debería figurar en la expresión [3.1.2] debería ser la diversidad calculada en el tiempo t , que podría denotarse como D_t . Si la medida de diversidad escogida es la diversidad de Weitzman, al estar ésta basada en distancias genéticas, las distancias al cabo del tiempo t entre los pares de unidades de un subconjunto S no tienen por qué ser iguales a las encontradas en el tiempo actual. Lo mismo sería aplicable a las probabilidades de extinción por generación. Se asume que éstas no varían con el paso de las generaciones, pese a que su variación puede ser importante, más aún cuando se calculan utilizando criterios sociopolíticos, económicos y culturales. Cuanto más contextual sea la información utilizada, más se agudizará este efecto.

DE_T es una función de (P_1, \dots, P_N) . Se define entonces la *diversidad marginal* para la especie i , $i = 1, \dots, N$, en la generación T como

$$[3.1.4] \quad DM_T(i) = -\frac{\partial DE_T}{\partial P_i}(P_i).$$

Obsérvese que la diversidad marginal será una cantidad positiva, ya que el valor de la diferencial será negativo, cosa intuitivamente clara, puesto que cuanto más riesgo de desaparecer tiene una población menor será en término medio la diversidad al cabo de un cierto número de generaciones.

Otro indicador interesante es la elasticidad de la diversidad, o potencial de conservación de cada población. La elasticidad es un concepto frecuente y sobradamente conocido en el ámbito de la economía. En la mayoría de los casos, compara las variaciones relativas en la cantidad de demanda con respecto a la variación relativa en los precios, obteniendo así una medida de la capacidad de respuesta o sensibilidad de la demanda respecto a cambios en los precios. Así, si por ejemplo, el precio de un bien se incrementa en un 2% y su demanda decrece en un 5%, la elasticidad de la demanda será de -2.5 . Normalmente la elasticidad de la demanda es negativa, ya que un incremento de precios suele tener como consecuencia una reducción en la cantidad de demanda. Si la elasticidad es menor que -1 , la demanda se dice elástica respecto a los precios, mientras que si está entre -1 y 0 , la demanda se dice inelástica. Si la elasticidad es igual a 0 , la demanda es completamente inelástica, lo que significa que el vendedor puede incrementar el precio del producto todo cuanto quiera y la demanda no decrecerá. El extremo opuesto es una demanda perfectamente elástica. En esta situación, la elasticidad se aproxima a $-\infty$, y el vendedor puede vender prácticamente todo lo que quiera al precio de mercado actual. Elevar los precios haría desaparecer prácticamente la demanda, y reducirlos sería contraproducente, puesto que la demanda es casi ilimitada con los precios actuales. La mayoría de las situaciones reales se sitúan en un término medio de estos dos extremos, y los vendedores han de adecuar los precios para optimizar los beneficios.

La fórmula general para la elasticidad de la demanda respecto a los precios es $E = ((\Delta Q_D / Q_D) / (\Delta P / P))$, donde P es el precio actual de un

determinado bien, Q_D la cantidad demandada a ese precio, ΔP un pequeño cambio en el precio actual y ΔQ_D un pequeño cambio en la cantidad demandada. Esta expresión se puede reordenar como $E = (\Delta Q_D / \Delta P)(P / Q_D)$. Cuando los cambios no se miden en un período corto de tiempo, se promedian los valores iniciales y finales de Q_D y P . Para aproximar la elasticidad en un punto concreto, se pueden tomar intervalos pequeños en torno a los valores de precio y demanda en ese punto y calcular diferencias, si bien se puede calcular exactamente si hay posibilidad de diferenciación.

El concepto de elasticidad es aplicable a muchas otras variables económicas, como oferta en lugar de demanda, o en contextos no económicos, como el de la diversidad genética. Si se establece una analogía entre la cantidad de demanda y la diversidad esperada por un lado, y entre el precio y la probabilidad de extinción por el otro, se puede definir la *elasticidad de la diversidad*, o, como fue renombrada por Weitzman (1993), *potencial de conservación* para la especie i , $i = 1, \dots, N$, en la generación T como

$$[3.1.5] \quad ED_T(i) = -\frac{\partial DED_T}{\partial P_i}(P_i) \cdot \frac{P_i}{DED_T},$$

donde como antes, se añade el signo negativo para manejar números positivos.

La utilidad de estos valores radica en que, como diferenciales que son, proporcionan información sobre el grado en que la variación en las probabilidades de extinción de las distintas poblaciones afecta a la diversidad esperada. La elasticidad es particularmente relevante cuando los costes de reducir la probabilidad de extinción de una población son directamente proporcionales a la propia probabilidad de extinción.

3.1.1.- Utilización de probabilidades de extinción genética en el cálculo de la diversidad marginal y elasticidad de la diversidad

Como se ha visto anteriormente, la buena gestión de un programa de conservación de la diversidad genética utilizando información del comportamiento futuro depende en última instancia de una correcta estimación de las probabilidades de extinción de las razas involucradas. Por esa razón es necesario contar con unas probabilidades obtenidas de forma científica, objetiva y rigurosa, cualidades que no cumplen las del ejemplo propuesto por Weitzman (1993). Reist-Marti *et al.* (2003) propusieron una estimación de las probabilidades de extinción basada en indicadores poblacionales, socioeconómicos, políticos y culturales, interesante en cuanto a que formaliza y cuantifica una serie de factores de indudable influencia en la extinción real de una raza, pero igualmente arbitraria desde el momento en que se seleccionan esos factores y no otros, y más aún cuando el peso que se le da a cada uno de ellos responde a consideraciones subjetivas. Bennewitz y Meuwissen (2005) utilizan exclusivamente información y metodología de genética de poblaciones para estimar parámetros relacionados con el crecimiento de las mismas y calcular con ellos probabilidades de extinción. Simianer (2005), por su parte, propone utilizar el número esperado de alelos. De forma similar, con un enfoque estrictamente genético, se propuso (García *et al.*, 2005) evitar la utilización del concepto de extinción *física*, en el sentido de considerar a una raza extinta cuando no exista ningún ejemplar representativo de ella vivo, sino el de *extinción genética*. El concepto es, además, tanto más apropiado cuanto que precisamente estamos calculando diversidades *genéticas* y la aplicación de estas herramientas está orientada a la conservación de recursos *genéticos*. Diremos que una raza está *genéticamente extinta* cuando todo locus en el genoma posee un alelo fijado en la población, es decir, cuando todos los individuos tienen exactamente el mismo genotipo, y éste es homocigoto en cualquier locus que se considere. De esta forma, y salvo fenómenos de mutación o inmigración, cualquier apareamiento entre dos individuos de la población dará lugar a un descendiente genéticamente idéntico a ellos y no habrá ninguna variación genética con respecto a la generación anterior. Como es imposible genotipar todo el genoma, se considerarán baterías de marcadores genéticos suficientemente representativos de la variabilidad genética de las razas estudiadas.

El cálculo de la probabilidad de extinción genética de una raza no es en absoluto trivial. Sea un conjunto de L locus, y para el locus i , sean $\{A_1^i, \dots, A_{n_i}^i\}$ sus alelos, $i = 1, \dots, L$. Para cada A_j^i , donde $j = 1, \dots, n_i$ e $i = 1, \dots, L$, sea p_j^i la frecuencia poblacional del alelo A_j^i en la generación t . Entonces la probabilidad de extinción de una raza en la t -ésima generación será igual a

$$[3.1.6] \quad P_{ext}(t) = \prod_{i=1}^L P\left[\bigcup_{j=1}^{n_i} (p_j^i = 1)\right] = \prod_{i=1}^L \left[\sum_{j=1}^{n_i} P[p_j^i = 1]\right],$$

ya que los sucesos son disjuntos (si un alelo se fija ningún otro puede hacerlo).

Ahora bien, la evolución de las frecuencias génicas (alélicas) a lo largo de las generaciones es un proceso complicado que depende de muy diversos factores, y cuyo estudio estadístico por medio de procesos de difusión permitirá llegar a expresiones concretas para $P[p_j^i = 1]$.

3.1.1.1.- Variación en el tiempo de las frecuencias alélicas.

Procesos de difusión.

Los diversos factores que influyen en la evolución con el tiempo de las frecuencias génicas se pueden clasificar dentro de tres grandes grupos (Crow y Kimura, 1970). El primero lo forman factores causantes de cambios directos, como mutación y migración con tasas constantes, o selección con intensidad constante. Este grupo de factores, denominado por Wright (1949) *presiones evolutivas sistemáticas*, provoca cambios deterministas en las frecuencias alélicas, así que nos centraremos en el segundo grupo de factores. Este segundo grupo está constituido por factores que producen fluctuaciones aleatorias en las frecuencias alélicas. De entre ellos se pueden distinguir a su vez dos tipos. Uno es el muestreo aleatorio de gametos en una población finita, es decir, la deriva genética, quizás el más importante, en especial en poblaciones de tamaño reducido. El otro es la fluctuación aleatoria en las

presiones sistemáticas, de las que la fluctuación en la intensidad de selección es quizá la más importante. Hay un tercer grupo de factores consistente en sucesos, como reordenamientos cromosómicos, duplicación o deficiencias en una secuencia de nucleótidos, o poliploidismo, que pueden ocurrir una vez a lo largo de la historia de una especie, y no se suelen tener en cuenta en la práctica para ser tratados sistemáticamente. En nuestro caso nos centraremos en el estudio de la influencia del segundo grupo de factores.

Un enfoque matemático que ha resultado muy eficaz en la modelización de la dinámica exacta de los procesos aleatorios que siguen las frecuencias génicas es el de las aproximaciones por procesos de difusión.

Sea una v.a. continua x_t , dependiente de un tiempo continuo t , y sea $\delta_x = x_{t+\delta_t} - x_t$, el cambio en x_t en un tiempo infinitesimal δ_t . Si x_t es un proceso de Markov y δ_t satisface que

$$\begin{aligned}
 E(\delta_x \mid x_t = x) &= M_{\delta_x}(x) \delta_t + o(\delta_t) \\
 \sigma^2(\delta_x \mid x_t = x) &= V_{\delta_x}(x) \delta_t + o(\delta_t) \\
 E(|\delta_x|^k) &= o(\delta_t) \quad \forall k \geq 3
 \end{aligned}
 \tag{3.1.7}$$

entonces se dice que x_t sigue un *proceso de difusión*.

En la práctica, los procesos que se van a considerar son todos markovianos (Crow y Kimura, 1970; Feller, 1951). Sin embargo, el hecho de que las variables que se manejan son en realidad discretas, así como el tiempo, que se mide en generaciones, hace que lo que se utilice sean aproximaciones por continuidad gracias a un reescalamiento de la variable y el tiempo, y estas aproximaciones sí siguen procesos de difusión (Lynch y Walsh, 2000). De hecho esta simplificación por el modelo de continuidad modela mejor la realidad que otras asunciones simplificadoras (Feller, 1951).

Las cantidades $M_{\delta_x}(x)$ y $V_{\delta_x}(x)$ que aparecen en [3.1.7] se conocen como *media infinitesimal* y la *varianza infinitesimal*, y se definen formalmente como

$$[3.1.8] \quad \begin{aligned} M_{\delta_x}(x) &= \lim_{\delta \rightarrow 0} \frac{E(x_{t+\delta} - x_t | x_t = x)}{\delta} \\ V_{\delta_x}(x) &= \lim_{\delta \rightarrow 0} \frac{E[(x_{t+\delta} - x_t)^2 | x_t = x]}{\delta} \end{aligned}$$

3.1.1.2.- Ecuaciones de Kolmogorov.

Las ecuaciones fundamentales que estudian el comportamiento de un proceso de difusión son las llamadas *ecuaciones de Kolmogorov* (Kolmogorov, 1931). Si x_t es la frecuencia de un determinado alelo segregando en una población en un tiempo t , y p_0 es la frecuencia en el tiempo inicial $t = 0$, entonces se tiene que la función de densidad de x_t , $\Phi(p_0, x; t)$ verifica la siguiente ecuación, la *ecuación de avance de Kolmogorov*:

$$[3.1.9] \quad \frac{\partial \Phi(p_0, x; t)}{\partial t} = \frac{1}{2} \frac{\partial^2}{\partial x^2} [V_{\delta_x}(x) \Phi(p_0, x; t)] - \frac{\partial}{\partial x} [M_{\delta_x}(x) \Phi(p_0, x; t)],$$

donde $0 < x < 1$. Es decir, la ecuación [3.1.9] no es válida en la frontera del espacio paramétrico, las denominadas *clases terminales* (Crow y Kimura, 1970).

La ecuación [3.1.9] es una aplicación al caso que nos ocupa de la *ecuación Fokker-Planck*, como se la conoce en ámbitos de la física.

Aunque el tiempo en la ecuación [3.1.9] es una variable continua, en la práctica puede considerarse en unidades de generaciones. En esta ecuación la frecuencia génica en la generación t es la variable aleatoria y la frecuencia inicial, p_0 , es un parámetro fijo. Si, en cambio, consideramos el proceso de cambio en las frecuencias alélicas en sentido contrario, es decir, hacia atrás en el tiempo, de forma que la frecuencia alélica x en la generación t sea una

cantidad fija, y la frecuencia inicial la variable aleatoria, se puede probar (Kolmogorov, 1931) que

$$[3.1.10] \quad \frac{\partial \Phi(p_0, x; t)}{\partial t} = \frac{1}{2} V_{\delta_x}(x) \frac{\partial^2}{\partial p^2} \Phi(p_0, x; t) + M_{\delta_x}(x) \frac{\partial}{\partial p} \Phi(p_0, x; t),$$

en la que se conoce como *ecuación de retroceso de Kolmogorov*.

Para conocer la probabilidad de fijación de un alelo al cabo de un determinado número de generaciones utilizaremos únicamente la ecuación de avance, como veremos a continuación.

3.1.1.3.- Probabilidad de extinción de un alelo al cabo de t generaciones.

En la aplicación de la ecuación [3.1.9] en la genética de poblaciones resulta conveniente tener en cuenta que la cantidad

$$[3.1.11] \quad P(x, t) := -\frac{1}{2} \frac{\partial}{\partial x} [V_{\delta_x}(x) \Phi(p_0, x; t)] + M_{\delta_x}(x) \Phi(p_0, x; t) \quad \forall 0 < x < 1$$

representa el flujo neto de masa de probabilidad a través de x en el tiempo t (Crow y Kimura, 1970).

El objetivo es calcular la probabilidad de que un alelo se fije al cabo de t generaciones, es decir, la probabilidad de que su frecuencia se iguale a uno en el transcurso de esas t generaciones. En primer lugar, asumiremos el caso especial, pero en absoluto restrictivo, en el que el cambio en las frecuencias de las clases terminales se debe enteramente a la entrada de flujo de probabilidad procedente de las *clases no fijas* ($0 < x < 1$,). Si denotamos $f(p_0, 1, t)$ la probabilidad que se está buscando, se tiene (Crow y Kimura, 1970) que

$$[3.1.12] \quad \frac{df(p_0, 1, t)}{dt} = P(1, t),$$

donde

$$[3.1.13] \quad P(1,t) = -\lim_{x \rightarrow 0} \left\{ \frac{1}{2} \frac{\partial}{\partial x} [V_{\delta_x}(x) \Phi(p_0, x; t)] + M_{\delta_x}(x) \Phi(p_0, x; t) \right\},$$

con lo que no hay más que integrar en [3.1.12] para obtener el resultado deseado.

3.1.1.4.- *Solución en poblaciones sometidas únicamente a deriva genética.*

La situación en la que la única fuente de fluctuación en las frecuencias génicas de una población es la debida al muestreo aleatorio de gametos en los procesos de apareamiento es, por sencilla, la más estudiada en diversos ámbitos de la genómica, desde genética de poblaciones hasta análisis de ligamiento o desequilibrio de ligamiento. Sin embargo, no por sencilla deja de aproximar razonablemente bien el comportamiento de muchas poblaciones naturales, por lo que es una hipótesis de partida con un peso cualitativo importante. En el caso que nos ocupa, la ventaja que ofrece el asumir una población en estas condiciones es la de proporcionar unas expresiones sencillas para $M_{\delta_x}(x)$ y $V_{\delta_x}(x)$, que permitirán resolver la ecuación de Kolmogorov de forma explícita y a su vez obtener una expresión concreta y fácilmente computable para $f(1,t)$. Se puede probar (Crow y Kimura, 1970, Karlin and Taylor, 1981) que en este caso se tiene que

$$[3.1.14] \quad M_{\delta_x}(x) = 0 \quad \text{y} \quad V_{\delta_x}(x) = \frac{x(1-x)}{2N_e},$$

siendo N_e el tamaño efectivo de la población, y que en adelante se asumirá por simplicidad igual a N , el tamaño poblacional.

De esta forma la ecuación [3.1.9] se convierte en

$$[3.1.15] \quad \frac{\partial \Phi(p_0, x; t)}{\partial t} = \frac{1}{4N} \frac{\partial^2}{\partial x^2} [x(1-x)\Phi(p_0, x; t)].$$

La solución explícita de [3.1.15] se obtuvo por Kimura (1955) en términos de la función hipergeométrica, y toma la forma siguiente:

$$[3.1.16] \quad \Phi(p_0, x; t) = \sum_{i=1}^{\infty} p_0(1-p_0)i(i+1)(2i+1) \cdot F(1-i, i+2, 2, p_0)F(1-i, i+2, 2, x)e^{\frac{-i(i+1)t}{4N}}$$

donde $F(\cdot, \cdot, \cdot, \cdot)$ es la función hipergeométrica, dada por

$$[3.1.17] \quad F(a, b, c, x) := 1 + \frac{ab}{1!c}x + \frac{a(a+1)b(b+1)}{2!c(c+1)}x^2 + \frac{a(a+1)(a+2)b(b+1)(b+2)}{3!c(c+1)(c+2)}x^3 + \dots$$

La solución [3.1.16] puede también expresarse (Korn and Korn, 1968) por medio de los *polinomios de Gegenbauer*, definidos como

$$[3.1.18] \quad T_{i-1}^1(z) = \frac{i(i+1)}{2} F\left(i+2, 1-i, 2, \frac{1-z}{2}\right)$$

de la siguiente forma:

$$[3.1.19] \quad \Phi(p_0, x; t) = \sum_{i=1}^{\infty} \frac{(2i+1)(1-r^2)}{i(i+1)} T_{i-1}^1(r) T_{i-1}^1(z) e^{\frac{-i(i+1)t}{4N}},$$

donde $r = 1 - 2p_0$ y $z = 1 - 2x$.

Por otra parte, se tiene, por [3.1.12] que

$$[3.1.30] \quad \frac{df(p_0, 1, t)}{dt} = \frac{1}{2} \Phi(p_0, 1, t) \frac{1}{2N} \frac{N}{N_e},$$

que bajo la asunción anterior es igual a $\frac{1}{4N} \Phi(p_0, 1, t)$.

Finalmente, se llega a que

$$[3.1.31] \quad \begin{aligned} f(p_0, 1, t) &= \frac{1}{4N} \int_0^t \Phi(p_0, 1, \tau) d\tau \\ &= p_0 + \frac{1}{2} \sum_{i=1}^{\infty} \frac{(2i+1)(1-r^2)}{i(i+1)} T_{i-1}^1(r) (-1)^i e^{\frac{-i(i+1)t}{4N}} \\ &= p_0 + \sum_{i=1}^{\infty} (2i+1) p_0 (1-p_0) F(i+2, 1-i, 2, p_0) (-1)^i e^{\frac{-i(i+1)t}{4N}} \end{aligned}$$

Esta cantidad es facilmente computable por un programa informático, puesto que la serie converge en pocas iteraciones y la serie que hay implícita en la definición de F en este caso concreto, con esos parámetros, deja de ser una serie para convertirse en una suma finita.

Por tanto, para los L loci en las condiciones previas a [3.1.6], si

$\{p_j^i(0)\}_{j=1, \dots, n_i}^{i=1, \dots, L}$ son las frecuencias iniciales de los alelos que los forman, se tiene que

$$[3.1.32] \quad P_{ext}(t) = \prod_{i=1}^L \left[\sum_{j=1}^{n_i} f(p_j^i(0), 1, t) \right].$$

3.2.- Minimización del parentesco medio y métodos core set

Como se ha visto anteriormente, la diversidad de Weitzman sólo contempla en su definición original como tal la variabilidad entre poblaciones, aspecto por el cual su utilidad ha sido discutida (Caballero y Toro, 2002; Eding *et al.*, 2002; Bruford, 2004), y también ha quedado patente que no es trivial extender la medida de forma que incorpore información intrapoblacional.

Eding y Meuwissen (2001) y Toro *et al.* (2002) propusieron sendos métodos para calcular parentescos entre individuos y entre poblaciones que posteriormente utilizaron para definir medidas de diversidad basadas en ellos que utilizan activamente la información genética dentro de poblaciones (Caballero y Toro, 2002; Eding *et al.*, 2002). Estas metodologías han sido utilizadas en diversas publicaciones en los últimos años, bien como aplicaciones en casos prácticos de determinación de prioridades en conjuntos de razas para la conservación de los recursos genéticos globales, bien en artículos de carácter metodológico en los que se profundiza en los aspectos técnicos (Bennewitz y Meuwissen, 2004; Fabuel *et al.*, 2004; Mateus *et al.*, 2004; Bennewitz *et al.*, 2006; European Cattle Genetic Diversity Consortium, 2006; Hayes *et al.*, 2006; Tapio *et al.*, 2006).

Utilizar el parentesco medio de una población, \bar{f} , parece una forma natural de medir su diversidad genética por varias razones. En primer lugar, porque su opuesto $1 - \bar{f}$, es, bajo condiciones de apareamiento aleatorio, la heterocigosis media, históricamente considerada una buena medida de diversidad. Además, \bar{f} está estrechamente relacionado con otros conceptos ampliamente utilizados en la genética de poblaciones, como la endogamia y el tamaño efectivo. De hecho, minimizar el parentesco medio es equivalente a

maximizar el tamaño efectivo (Caballero y Toro, 2002). Por otro lado, buena parte de las distancias genéticas más habituales (entre ellas las más utilizadas en los estudios de diversidad genética, las de Nei y la de Reynolds, Baumung *et al.*, 2004) pueden reescribirse en términos de parentescos dentro y entre poblaciones. Finalmente, en un modelo de sólo deriva (aplicable porque la escala temporal de la formación de las razas habitualmente estudiadas es relativamente corta, con lo que no se espera que la mutación tenga un impacto importante) la diversidad en términos de parentesco medio es proporcional a la máxima varianza genética que se puede recuperar en una población formada por apareamiento aleatorio uniendo todas las subpoblaciones, ya que en esas circunstancias, $\sigma_w^2 = (1 - \bar{f})\sigma_a^2$, donde σ_a^2 es la varianza genética total original (Falconer y Mackay, 1996; Eding *et al.*, 2002).

3.2.1.- Contribuciones al parentesco medio y descomposición en componentes entre y dentro de poblaciones

Se propone utilizar como medida de diversidad uno menos el parentesco medio del conjunto de poblaciones, entendido éste como la media de los parentescos dentro y entre poblaciones. Supongamos por ahora que todos los parentescos son conocidos genealógicamente, i.e., que estamos tratando con probabilidades de IBD. Para un conjunto P de n poblaciones, tendremos una matriz de parentescos M de tamaño $n \times n$, con $M_{ij} = f_{ij}$, el parentesco medio entre las poblaciones i y j , calculado como la media de todos los parentescos entre un individuo de i y uno de j . M_{ii} será, por tanto el parentesco medio de la población i , calculado como la media de los parentescos entre pares de individuos de i (no confundir con la media de los parentescos de cada individuo consigo mismo, el llamado autoparentesco).

Si N_i es el tamaño de la población i ($i = 1, \dots, n$), y N_T el tamaño total de la población conjunta ($N_T = N_1 + \dots + N_n$), el parentesco medio \bar{f} será igual a

$$\bar{f} = \frac{\sum_{i=1}^n \sum_{j=1}^n f_{ij} N_i N_j}{N_T^2},$$

o, en notación matricial, $\bar{f} = w^t M w$, con $w = (w_1, \dots, w_n)^t$, y w_i la fracción poblacional correspondiente a la población i , N_i/N_T , es decir, el peso que le corresponde a la población i en el conjunto de las n poblaciones.

Teniendo en cuenta que la distancia mínima de Nei (Nei, 1987) entre las poblaciones i y j se puede escribir como

$$D_m(i, j) = \frac{f_{ii} + f_{jj}}{2} - f_{ij},$$

el parentesco medio se puede reescribir como (Caballero y Toro, 2002)

$$\bar{f} = \sum_{i=1}^n w_i \left[f_{ii} - \sum_{j=1}^n w_j D_m(i, j) \right].$$

Esta expresión resulta interesante porque permite ver el parentesco medio como una media ponderada de factores en los que intervienen para cada población i , por un lado, las diversidades *dentro* de poblaciones, f_{ii} , y las diversidades *entre* poblaciones, en la forma de la media de las distancias de todas las poblaciones con respecto a la población i . Así, cada miembro del sumatorio externo es la contribución que cada población hace a la diversidad total. Del cálculo de esta expresión se pueden extraer para cada población dos parámetros de interés para el estudio de las prioridades de conservación entre las distintas poblaciones (Caballero y Toro, 2002). Uno sería la propia contribución a la diversidad de la población,

$$c_i = w_i \left[f_{ii} - \sum_{j=1}^n w_j D_m(i, j) \right],$$

o, en términos relativos, c_i/\bar{f} . El otro sería la pérdida relativa de diversidad por la eliminación de la población i , es decir, la relación entre la diferencia existente entre la diversidad total y la diversidad recalculada sin la población i , y la diversidad total, es decir,

$$l_i = \frac{\bar{f} - \bar{f}_i}{1 - \bar{f}},$$

con \bar{f}_i la diversidad calculada sin la población i .

3.2.2.- Contribuciones óptimas para minimizar el parentesco medio. Método del Core Set

Este último parámetro, l_i , presenta un problema, y es que al ser \bar{f} un valor medio, puede darse el caso de que al eliminar una población, el nuevo valor medio global sea menor que el anterior, es decir, que el parentesco medio disminuya, y por tanto se tenga un aumento de la diversidad, lo que contradice no sólo las normas propuestas por Weitzman (1992) para una buena medida de diversidad, sino el sentido común.

Por ejemplo (Eding *et al.*, 2002), si se consideran tres poblaciones de igual tamaño tales que su matriz de parentescos medios sea

$$M = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

(es decir, la primera no tiene relación alguna con la segunda y la tercera, y estas dos últimas son idénticas), entonces el parentesco medio \bar{f} será 5/9. Si se eliminara la tercera población, la nueva matriz sería

$$M_{|3} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

y el nuevo parentesco medio \bar{f}_3 sería igual a $2/4$, que es menor que $5/9$, con lo que la diversidad habría aumentado al prescindir de una de las poblaciones.

Esto sucede al trabajar con medias y tener poblaciones que comparten mucha diversidad (en este caso, la misma población duplicada), es decir, existe un solapamiento genético que hace prescindible una buena parte de la diversidad aportada por una de las poblaciones (toda en este caso).

Este problema se evitaría basando la definición de diversidad no en el parentesco medio de la población conjunta tal y como está definida, sino en el de un subconjunto “núcleo” (*core set*) S en el que las poblaciones estuvieran representadas en unas proporciones tales que el solapamiento genético fuera mínimo. Esto se conseguiría minimizando el parentesco medio dentro de ese núcleo (Eding *et al.*, 2002).

En la práctica, lo que proponen Eding *et al.* (2002) es variar las proporciones en las que las distintas poblaciones entrarían a un modelo de apareamiento aleatorio entre sí. El parentesco medio calculado anteriormente, \bar{f} , es que le correspondería a una metapoblación en la que se juntaran todas las subpoblaciones de la población global y los individuos se aparearan aleatoriamente. Lógicamente cada supoblación i aportaría a esta unificación una proporción igual a $w_i = N_i/N_T$, que es la proporción en la que está representada en la población conjunta actual. Lo que se busca ahora es un nuevo vector de pesos w de tal forma que el parentesco medio calculado con ellos sea mínimo. Caballero y Toro (2002) plantearon este mismo marco de trabajo no como solución al problema de pérdidas positivas (i.e., ganancias) de diversidad al eliminar razas, sino en el contexto de aportaciones reales de las distintas razas a un pool de conservación *in vivo* o a un programa de crioconservación, buscando formar una población sintética o un banco de germoplasma. Con matices que se analizarán más adelante, el punto de partida es el mismo:

$$\min_w \bar{f}(w) = \sum_{i=1}^n w_i \left[f_{ii} - \sum_{j=1}^n w_j D_m(i, j) \right] = w^t M w, \text{ s.a. } w_i \geq 0 \quad \forall i = 1, \dots, n \text{ y } \sum_{i=1}^n w_i = 1.$$

Caballero y Toro (2002) solucionan este problema mediante métodos iterativos, y añaden los w_i^{\min} , las proporciones en que las distintas poblaciones contribuirían al pool, a los parámetros anteriores, c_i y l_i , como una información adicional para estudiar la determinación de prioridades de conservación.

La solución planteada por Eding *et al.* (2002) al mismo problema es analítica, basada en los multiplicadores de Lagrange. Así, se establece una función $\Phi(w, \lambda) = w^t M w - \lambda(w^t 1_n - 1)$, donde 1_n es un vector n -dimensional de unos, se calculan sus diferenciales parciales y se llega finalmente a la solución

$$w^{\min} = \frac{M^{-1} 1_n}{1_n^t M^{-1} 1_n}.$$

De esta forma, se tiene que

$$\bar{f}^{\min} = (w^{\min})^t M w^{\min} = \frac{1}{1_n^t M^{-1} 1_n},$$

y la diversidad máxima en el pool o conjunto núcleo será $Div(S) = 1 - \bar{f}^{\min}$. Así, además de aportar los valores w_i^{\min} , se pueden también recalcular las contribuciones a la diversidad maximizada utilizando la descomposición anterior en diversidad dentro y entre, además de las pérdidas por la desaparición de una de las poblaciones, que en este caso ya no serían positivas en ningún caso.

Puede darse el caso de que existan contribuciones negativas en la solución w^{\min} , debido a errores muestrales en la estimación de los parentescos o a efectos poblacionales como que los parentescos medios de una población con respecto a otra(s) sean mayores que el parentesco medio dentro de ella.

En tal caso, se propone (Eding *et al.*, 2002) igualar la más negativa de esas contribuciones a cero y recalcular w^{\min} sin la población correspondiente, procediendo análogamente hasta que todos los w_i^{\min} sean no negativos. Esto aleja la solución final de la óptima encontrada en el primer paso, pero aún así parece que es la mejor posible para las restricciones impuestas.

En realidad, la solución óptima pasaría por hacer cero una a una cada una de las contribuciones negativas, recalcular las contribuciones del resto de poblaciones y ver si en la nueva solución todas las contribuciones son positivas, en cuyo caso se estaría ante un óptimo restringido. Si no fuera así, se procedería a hacer cero las contribuciones negativas de dos en dos, luego de tres en tres, y así sucesivamente hasta que se diera con una nueva solución en la que las contribuciones restantes fueran todas positivas (Walsh, 1979; Cooper, 1981; Pike, 1986). Además, los autores no garantizan que el óptimo que se encuentra sea un mínimo, y es importante que así sea, ya que de encontrarse en su lugar un máximo se estaría consiguiendo el objetivo opuesto al deseado, i.e., minimizar la diversidad en lugar de maximizarla. Sin embargo, como se verá más adelante, se puede probar que en las condiciones teóricas en las que se realiza la optimización, se llega efectivamente a un mínimo. Por otra parte, Eding *et al.* (2002) afirman que al minimizar el parentesco medio en el conjunto núcleo se evita que la desaparición de una población haga aumentar la diversidad, o, viceversa, que añadir una nueva población haga que la diversidad disminuya. Si bien esto es cierto cuando se alcanza el óptimo en el primer paso, ya que eliminar una población con una contribución no nula implica que el nuevo mínimo es “menos” mínimo que el anterior, esto puede no ser así cuando se han realizado una serie de iteraciones encaminadas a eliminar contribuciones negativas, en cada una de las cuales el nuevo óptimo encontrado proporciona un valor de parentesco medio ligeramente superior al anterior.

Existe una diferencia más entre la metodología propuesta por Caballero y Toro (2002) y la de Eding *et al.* (2002), y es la relativa a la estimación de los parentescos moleculares que dan lugar a la matriz M . Si los parentescos fueran

conocidos a nivel de pedigrí, ambos procedimientos ofrecerían resultados más similares (no iguales puesto que el proceso de búsqueda del óptimo es distinto). Mientras que Caballero y Toro utilizan el parentesco molecular (i.e., la aplicación directa de la definición de Malecot (1948)), Eding *et al.* utilizan la estimación del parentesco genealógico obtenida corrigiendo el parentesco molecular por la probabilidad de IBS. Esto no sería necesario bajo la hipótesis de que los alelos en la población inicial fueran todos distintos, en cuyo caso el parentesco molecular es un buen estimador del parentesco genealógico, pero esa hipótesis no siempre es asumible. Para el cálculo del parentesco molecular se asume apareamiento aleatorio, de forma que se pueden utilizar frecuencias alélicas como se vio en el apartado 2.4. En otro caso, se puede recurrir a calcular los parentescos (moleculares) entre pares de individuos y promediar para obtener los parentescos medios. Caballero y Toro contemplan también la posibilidad de utilizar estimadores del parentesco genealógico, pero siguiendo los resultados de Toro *et al.* (2001), comentados en el apartado 2.4, desaconsejan su uso. Sin embargo, como prueban Eding *et al.*, el vector de contribuciones al conjunto núcleo, w^{\min} , es invariante frente a la probabilidad de identidad en estado cuando ésta se considera única para todos los loci.

Las distintas formas de estimar los parentescos genealógicos presentadas por Eding y Meuwissen (2003) resultan en distintas soluciones para el vector de contribuciones. El modelo *WLM* ofrece un buen compromiso entre la precisión de la estimación de los parentescos y la del vector de contribuciones. El *WLMM*, en cambio, proporciona una correlación algo más baja con los parentescos reales, y un óptimo algo peor en el proceso de maximización de la diversidad. Sin embargo, las contribuciones obtenidas con este modelo correlacionan más que las de ningún otro con las reales, y es el que menos contribuciones nulas proporciona. De hecho se puede ajustar de forma que la menor contribución sea cero, y el resto sean superiores a esa, con lo que se minimiza el número de poblaciones que quedarían excluidas, además de poder establecer un ordenamiento de todas las poblaciones sin excepción en función de su contribución a la diversidad del conjunto núcleo.

3.2.3.- Condición de mínimo de la solución del proceso de optimización

Si bien Eding *et al.* (2002) asumen que su método de optimización es en realidad uno de minimización, lo cierto es que en principio la solución encontrada no tiene por qué ser un mínimo, o, al menos, no se ofrece una prueba de que así sea. Se probará en esta sección que en las condiciones en las que se realiza el proceso de optimización, con una matriz de parentescos, la solución a la que se llega es, en efecto, un mínimo.

En la situación más sencilla, cuando sólo intervienen dos poblaciones, la forma cuadrática $w^t M w$ se puede expandir de la forma siguiente:

$$w^t M w = \begin{pmatrix} w_1 & w_2 \end{pmatrix} \begin{pmatrix} m_1 & m_2 \\ m_3 & m_4 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = m_1 w_1^2 + (m_2 + m_3) w_1 w_2 + m_4 w_2^2$$

Como función de w , $w^t M w$ se puede representar gráficamente, e imponiendo la restricción $c^t 1_n = 1$, el óptimo se puede ver fácilmente en la figura (Fig. 3.2.1). Analíticamente, $c^t 1_n = 1$ se reduce a $c_2 = c_1 - 1$, y la optimización restringida en las dos variables se reduce a encontrar óptimos en la función de una variable

$$\begin{aligned} f(c_1) &= m_1 c_1^2 + (m_2 + m_3) c_1 (1 - c_1) + m_4 (1 - c_1)^2 = \\ &= (m_1 - m_2 - m_3 + m_4) c_1^2 + (m_2 + m_3 - 2m_4) c_1 + m_4 \end{aligned}$$

Ahora bien, el óptimo de $f(c_1)$ puede ser un máximo o un mínimo. El signo de la segunda derivada, que es igual a $2(m_1 - m_2 - m_3 + m_4)$, será positivo o negativo dependiendo de los valores de los elementos de la matriz de parentescos. Así, se tendrá un mínimo cuando $m_1 + m_4 > m_2 + m_3$, o equivalentemente, cuando el parentesco medio dentro de poblaciones sea mayor que el parentesco medio entre poblaciones, como de hecho suele suceder. Pero si se diera la situación opuesta, la solución encontrada sería un máximo y la diversidad en el conjunto núcleo se estaría minimizando.

La figura 3.2.1a muestra la función para un ejemplo en el que los parentescos para las poblaciones 1 y 2 y el parentesco entre ellas son, respectivamente 0.2834, 0.3889 y 0.5278. La parte superior de la figura es una proyección en el plano c_1c_2 , de forma que las coordenadas del valor óptimo se pueden aproximar gráficamente. Analíticamente, este valor se obtendría como $M^{-1}1_n / 1_n' M^{-1}1_n$, y sería igual a $c_1 = 0.3624$, $c_2 = 0.6376$. Estos valores se corresponden con los observados gráficamente, pero como se puede ver en la figura, el óptimo en ese caso es un máximo, con lo que el resultado no es válido para el propósito original de conservación de la diversidad.

La figura 3.2.1b muestra la situación contraria, en la que los parentescos intrapoblacionales son 0.3332 y 0.3469, y el parentesco entre poblaciones es igual a 0.2697. La condición de mínimo se verifica, y los valores teóricos $c_1 = 0.5487$ y $c_2 = 0.4513$ se corresponden efectivamente con los de un mínimo para la función restringida, como se puede observar en la figura.

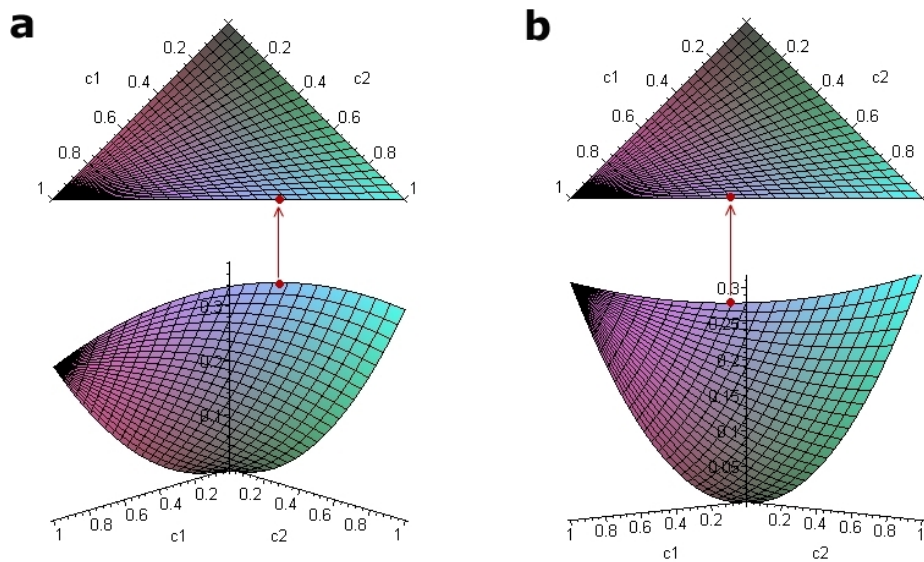


Fig. 3.2.1: ejemplos en los que se verifican las diferentes condiciones para encontrar un máximo (a) y un mínimo (b). La parte superior de las figuras muestra la proyección en el plano c_1c_2 , de forma que las coordenadas de las soluciones se pueden aproximar visualmente.

Para el caso general n -dimensional, se necesitan encontrar análogamente condiciones suficientes para tener máximos o mínimos. Pike (1986), a partir de un teorema general de Avriel (1976) propuso unas condiciones en la forma siguiente:

Sean $y(x)$ y $f_i(x) = 0, i = 1, 2, \dots, m$ funciones reales doblemente diferenciables de forma continua, donde $x \in \mathbb{R}^n$. Si existen vectores $x^* \in \mathbb{R}^n$ and $\lambda^* \in \mathbb{R}^m$, tales que $L_i(x^*, \lambda^*) = 0, i = 1, \dots, m$, donde L_i es la diferencial parcial de la función lagrangiana $L(x, \lambda) = y(x) + \sum_{j=1}^m \lambda_j f_j(x)$ con respecto a x_i , i.e., x^* es un óptimo de y sujeto a las restricciones impuestas por $f_i(x)$, y si

$$D_p = (-1)^m \begin{vmatrix} L_{11} & \dots & L_{p1} & f_{11} & \dots & f_{m1} \\ \vdots & & \vdots & \vdots & & \vdots \\ L_{1p} & \dots & L_{pp} & f_{1p} & \dots & f_{mp} \\ f_{11} & \dots & f_{1p} & 0 & \dots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ f_{m1} & \dots & f_{mp} & 0 & \dots & 0 \end{vmatrix} > 0$$

para $p = m+1, \dots, n$, donde L_{ij} y f_{ij} corresponden a $\frac{\partial L}{\partial x_i \partial x_j}(x^*, \lambda^*)$ y $\frac{\partial f_i}{\partial x_j}(x^*, \lambda^*)$,

respectivamente, entonces $y(x)$ tiene un mínimo local estricto en x^* tal que

$f_i(x^*) = 0, i = 1, 2, \dots, m$. Cambiando $(-1)^m$ por $(-1)^p$ en la expresión superior se obtiene una condición equivalente para un mínimo local estricto.

Como ejemplo sencillo de aplicación, en el caso bidimensional anterior, se llega, como era de esperar, a la misma condición sobre m_1, \dots, m_4 obtenida anteriormente:

Se quiere optimizar $y(c) = m_1 c_1^2 + (m_2 + m_3) c_1 c_2 + m_4 c_2^2$ sujeta a $f_1(c) = c_1 + c_2 - 1 = 0$. Entonces,

$$L(c, \lambda) = m_1 c_1^2 + (m_2 + m_3) c_1 c_2 + m_4 c_2^2 + \lambda_1 (c_1 + c_2 - 1),$$

$$L_1(c, \lambda) = 2m_1 c_1 + (m_2 + m_3) c_2 + \lambda_1$$

$$L_2(c, \lambda) = 2m_4c_2 + (m_2 + m_3)c_1 + \lambda_1$$

$$L_{11} = 2m_1, L_{12} = m_2 + m_3, L_{21} = m_2 + m_3, L_{22} = 2m_4$$

$$f_{11} = 1, f_{12} = 1$$

y

$$D_2 = (-1)^1 \begin{vmatrix} 2m_1 & m_2 + m_3 & 1 \\ m_2 + m_3 & 2m_4 & 1 \\ 1 & 1 & 0 \end{vmatrix} = 2(m_1 + m_4 - m_2 - m_3),$$

con lo que $D_2 > 0$, la condición para un mínimo, se alcanzará cuando $m_1 + m_4 > m_2 + m_3$, como se vio anteriormente. Si se usa $(-1)^2$ en lugar de $(-1)^1$ la condición para un máximo aparece también de nuevo como $m_1 + m_4 < m_2 + m_3$.

En el caso general, si se denota $M = (m_{ij})_{i,j=1,\dots,n}$, las funciones anteriores se reducen a

$$y(c) = c^t M c = \sum_{i=1}^n \sum_{j=1}^n m_{ij} c_i c_j = \sum_{i=1}^n m_{ii} c_i^2 + 2 \sum_{i=1}^n \sum_{j=i+1}^n m_{ij} c_i c_j$$

$$f_1(c) = c^t 1_n - 1 = \sum_{i=1}^n c_i - 1$$

$$L(c, \lambda) = \sum_{i=1}^n m_{ii} c_i^2 + 2 \sum_{i=1}^n \sum_{j=i+1}^n m_{ij} c_i c_j + \lambda \left(\sum_{i=1}^n c_i - 1 \right)$$

$$L_i(c, \lambda) = 2 \sum_{j=1}^n m_{ij} c_j + \lambda, i = 1, \dots, n$$

$$L_{ij}(c, \lambda) = 2m_{ij}, i, j = 1, \dots, n$$

$$f_{1i}(c) = 1, i = 1, \dots, n$$

Las condiciones suficientes para un mínimo son entonces

$$D_p = (-1)^p \begin{vmatrix} 2m_{11} & \dots & 2m_{p1} & 1 \\ \vdots & & \vdots & \vdots \\ 2m_{1p} & \dots & 2m_{pp} & 1 \\ 1 & \dots & 1 & 0 \end{vmatrix} > 0 \quad \forall p = 2, \dots, n.$$

Denotando por M_p la submatriz de M cuadrada simétrica formada por las p primeras filas y columnas, entonces

$$\begin{aligned} D_p &= (-1) \begin{vmatrix} 2M_p & 1_p \\ 1_p^t & 0 \end{vmatrix} = (-1) \begin{vmatrix} 2M_p & 0 \\ 1_p^t & -1_p^t (2M_p)^{-1} 1_p \end{vmatrix} = \\ &= (-1) 2^p \begin{vmatrix} M_p & 0 \\ 1_p^t & -\frac{2^{p-1}}{2^p} 1_p^t M_p^{-1} 1_p \end{vmatrix} = 2^{p-1} \begin{vmatrix} M_p & 1_p^t \\ 1_p^t & M_p^{-1} 1_p \end{vmatrix} \end{aligned}$$

Ahora bien, si M es una matriz de parentescos medios, es definida positiva, por lo que tanto M_p como M_p^{-1} son definidas positivas, y por tanto $|M_p| > 0$ y

$1_p^t M_p^{-1} 1_p > 0$. Así, salvo errores en la estimación de los parentescos, la solución encontrada siempre será un mínimo.

3.2.4.- Extensión del método Core Set para minimizar la varianza de un carácter cuantitativo

El método del Core Set de Eding *et al.* maximiza la variabilidad que se encontraría en la hipotética descendencia de una población sintética en la que las distintas subpoblaciones se cruzaran aleatoriamente en proporciones correspondientes a los vectores de contribuciones óptimas. Bennewitz y Meuwissen (2005) propusieron un método análogo que maximiza la varianza genética total de un hipotético carácter cuantitativo. Para un conjunto S de N poblaciones, esta varianza se puede escribir como

$$\text{var}(S) = \frac{1}{N} \sum_{i=1}^N [(1 + M_{ii}) - 2\bar{M}],$$

siendo \bar{M} la media de los valores de la matriz de parentescos M .

Si de nuevo consideramos un conjunto nuclear en el que se maximice esta varianza, S_{mvt} , y las contribuciones con las que cada población entraría a

formar parte de él, almacenadas en un vector w_{mvt} de tamaño N , la función a maximizar sería, en notación matricial,

$$\text{var}_{tot}(S_{mvt}) = 1 + w_{mvt}^t F - 2w_{mvt}^t M w_{mvt},$$

siendo $F = \text{diag}(M)$, el vector con los elementos de la diagonal de M , que contiene los parentescos dentro de poblaciones.

Con un procedimiento análogo al del Core Set original, se obtiene el vector de ponderaciones que maximiza la varianza, que viene dado por la expresión

$$w_{mvt} = \frac{1}{4} \left(M^{-1} F - \frac{1_N^t M^{-1} F - 4}{1_N^t M^{-1} 1_N} \cdot M^{-1} 1_N \right).$$

Las diferencias en las varianzas obtenidas con las contribuciones de los dos métodos de Core Set no son demasiado grandes, pero este segundo método da más peso a las diferencias entre poblaciones, aumentando las contribuciones de aquéllas que están más separadas. Para una comparación más detallada de ambos métodos, véanse Bennewitz y Meuwissen (2005) y Bennewitz *et al.* (2006).

3.3.- Contribuciones globales combinadas

Una forma sencilla de tener en cuenta tanto la diversidad entre poblaciones como dentro de ellas cuando se trata de estimar las contribuciones de cada una a la diversidad global es la propuesta por Ollivier y Foulley (2002, 2004, 2005). Se trata de crear una nueva contribución en la que se ponderen las contribuciones obtenidas por procedimientos que consideren variabilidad dentro y variabilidad entre, mediante una combinación lineal de estas cantidades.

Si se considera una medida de diversidad estrictamente entre poblaciones, como la de Weitzman, se tiene para cada población i su contribución parcial a la diversidad, según la expresión [3.1.1]. De forma análoga se puede proceder con una medida de diversidad dentro de poblaciones, como la heterocigosis esperada, y calcular la proporción de heterocigosis total que se perdería al eliminar la población i . Así,

$$CP_i^w = \frac{H_e(Q) - H_e(Q \setminus i)}{H_e(Q)},$$

donde $H_e(Q)$ es la heterocigosis esperada para el conjunto de todas las poblaciones, y $H_e(Q \setminus i)$ la correspondiente al mismo conjunto excluyendo los individuos de la población i . Denotando por CP_i^B la contribución resultante de [3.1.1], se tiene, pues, para cada población, una pareja de contribuciones a los dos componentes de la diversidad conjunta.

Si ahora se toma un factor $0 \leq \alpha \leq 1$, se puede definir una contribución agregada como

$$CP_i^A = \alpha CP_i^B + (1 - \alpha) CP_i^W .$$

El problema se reduce por tanto a una elección adecuada de ese factor. Esta elección no deja de ser un tanto arbitraria, aunque parece razonable que ya que el F_{ST} es la proporción de varianza genética total atribuible a las diferencias entre poblaciones, se utilice aquí como ponderación, tomando por tanto de forma natural $\alpha = F_{ST}$. Este sería el enfoque ideal cuando lo que se pretende es la creación de una metapoblación con el objetivo de hacer una mejora genética por selección (Ollivier y Foulley, 2002, 2005). Sin embargo, en un contexto más general, podría estar justificado aumentar el peso de la componente entre poblaciones (Chaiwong y Kinghorn, 1999; Piyasatian y Kinghorn, 2003) para explotar los beneficios derivados de la heterosis mediante cruzamientos, e incluso se puede llegar a dar hasta cinco veces más peso a las contribuciones “entre” que a las “dentro” (Chaiwong y Kinghorn, 1999; Ollivier y Foulley, 2002), con lo que α sería igual a 5/6. La arbitrariedad de la elección de la ponderación se convierte así en la principal ventaja que proporciona el concepto de contribución agregada, ya que aporta una gran flexibilidad al poder escoger la ponderación más adecuada a cada caso concreto según los objetivos que se persigan.

PARTE II

Conservación de la diversidad genética
de la raza de lidia

4.- Introducción a la raza de lidia

La raza bovina de lidia es la raza española con mayor trascendencia fuera de sus fronteras de origen. Su peculiar sistema de producción, junto con una demanda muy diversa de productos en función del tipo de festejo, ha dado lugar a una estructura de población característica, dividida en líneas o encastes, incluso en ganaderías dentro de encastes. Su expansión incluye países europeos como Francia, España y Portugal, y numerosos países de Centro y Sudamérica (Cañón *et al.*, 2008). El conocimiento de la existencia de estructuras en las poblaciones, puede ser de gran utilidad para clasificar a las ganaderías o a los denominados encastes, definir unidades de conservación genética y entender, identificar o corroborar procesos que han podido dar lugar a la variabilidad que actualmente podemos observar en la raza de lidia. Métodos demográficos junto con estudios del comportamiento de la variabilidad genética constituyen los principales elementos a tener en cuenta en la elaboración de programas de conservación.

La raza de lidia presenta una serie de particularidades que la alejan de la mayoría de las razas bovinas, destacando entre todas el hecho de que es la única cuyo criterio de selección se centra en caracteres de comportamiento, como son las diferentes respuestas de agresividad en los diferentes tipos de festejos (Silva *et al.*, 2002, 2006). Inicialmente no existía una cría propia para tal fin, utilizándose el ganado que destinado a la producción de carne mostraba mayor dificultad al manejo. Posteriormente, y debido a la particularidad del objetivo, además del arraigo de la lidia en los festejos populares y la notoriedad que adquirirían los ganaderos, surgió la cría específica de ganado para la agresividad en la lidia. Geográficamente esa cría se situó alrededor de las grandes cuencas fluviales dando lugar a las vacadas o castas fundacionales que han originado los actuales encastes a partir de un largo proceso de casi tres siglos de selección.

La proximidad geográfica entre las vacadas favorece, en principio, la homogeneidad genética o la continuidad genética entre los individuos de la raza de lidia. No obstante, los diferentes tipos de festejos y, por ello, diferentes objetivos de selección ha influido en la formación de subpoblaciones, llamadas encastes, cuyo comportamiento y morfología es diferente. Así, por ejemplo, en el estándar racial de la raza de lidia están admitidas una gran variedad de capas y de particularidades que en otras razas están muy restringidas (Boletín Oficial del Estado, 2001).

4.1.- Origen de la raza de lidia

4.1.1.- Origen filogenético

El antepasado más remoto del toro de lidia, al igual que el de la mayoría de los bóvidos actuales, es el uro (*Bos primigenius*), una especie de rumiante salvaje que los celtas llamaron “auroch” (de *aur*, salvaje; y *och*, toro) y en latín se conoció como “urus” (Zeuner, 1963; Grigson, 1978, 1980; Epstein y Mason, 1984; Payne, 1991). En la actualidad se reconocen tres subespecies de *Bos primigenius* según su localización geográfica, *Bos primigenius primigenius* en el noroeste de Eurasia, *Bos primigenius namadicus* en el sureste de Asia y *Bos primigenius opisthonomus* en África (Payne, 1970; Epstein y Mason, 1984), que son los antecesores de las razas bovinas actuales. Concretamente, del *Bos primigenius primigenius* se originaron los actuales taurinos (Epstein, 1971; Epstein y Mason, 1984), pertenecientes a la subespecie *Bos taurus taurus*, y del *Bos primigenius namadicus* los cebuinos, *Bos taurus indicus*. Esta bifurcación se produjo hace unos 100.000 años, mucho antes de que comenzara la domesticación (Manwell y Baker, 1980; Baker y Manwell, 1991; Bradley *et al.*, 1994, 1996; Loftus *et al.*, 1994).

Las primeras evidencias de la domesticación de los bovinos datan de hace 8400 años y fueron encontradas en un yacimiento del Neolítico en Anatolia (Turquía). Otros autores sitúan el origen de la domesticación hace 8000-10000 años en el sudeste de los Balcanes, basándose en las representaciones artísticas encontradas en restos de cerámica (Reed, 1977; Loftus *et al.*, 1994).

El uro sufrió varios procesos de domesticación en diferentes momentos y regiones del mundo. Mientras que en el subcontinente indio el bovino

doméstico que aparece pertenece a la subespecie *Bos taurus indicus*, tipo cebuino, bóvidos con joroba, en Europa el tipo de bóvido doméstico que aparece es el perteneciente a la subespecie *Bos taurus taurus*, tipo taurino, bóvidos sin joroba. Análisis realizados en razas bovinas de África, Europa y Asia evidenciaron una clara separación entre los taurinos y cebuinos, proporcionando mayor veracidad a la teoría del doble origen de la domesticación (MacHugh *et al.*, 1997). A partir de la domesticación, los movimientos humanos fueron acompañados de movimientos de los animales, influyendo estos procesos migratorios en la actual distribución de las razas bovinas.

En África ya existía una especie de Uro indígena (*Bos primigenius opisthonomus*), del que se ha discutido si era una forma africana o se originó a raíz de las primeras poblaciones que llegaron al continente (Epstein, 1971; Epstein y Mason, 1984; Payne, 1991), e incluso se plantea un tercer origen de domesticación independiente del ganado africano que explicaría las diferencias genéticas entre los taurinos africanos y europeos (Bradley *et al.*, 1996; Troy *et al.*, 2001).

La Península Ibérica, dada su proximidad a África, constituye una región que ha podido recibir influencias tanto de poblaciones bovinas provenientes del norte de Europa, como de otras provenientes de África. La llegada de los bóvidos a la Península, fechada en el año 7700 a. de C. en la costa oriental, se produce así por dos vías, la primera desde Europa atravesando los Pirineos, y la segunda a partir de las poblaciones originarias de Asia que atraviesan el istmo de Suez entrando en Egipto y expandiéndose hacia el norte, este y oeste de África hasta fusionarse con los cebuinos, entrando en la Península Ibérica a través del estrecho de Gibraltar, y encontrándose con los taurinos que entraron por lo Pirineos (Cymbron 1999; Beja-Pereira, 2003). A pesar de que tradicionalmente la influencia de las poblaciones africanas se asociaba a la llegada de los cartagineses primero (unos 200 años a.C.) y a la de los bereberes después, recientemente se han encontrado en la península restos de bovino de la edad del bronce (~ 3.800 años) cuyo origen es claramente africano (Anderung, 2005). La influencia

africana sobre algunas de las razas bovinas de la Península Ibérica ha sido corroborada recientemente con ADN mitocondrial, y en la raza de lidia aparece con una intensidad muy diferente entre las distintas ganaderías (Cymbron *et al.*, 1999; Miretti *et al.*, 2004; Cortés, 2008; Cortés *et al.*, 2007, 2008).

Existe cierto interés en demostrar que el toro de lidia es el representante más directo del extinguido uro, y que la bravura es un atributo heredado de la fiereza primitiva. Los ganaderos españoles crearon la raza de lidia con los mismos procedimientos zootécnicos que han sido aplicados en otras razas: el cruzamiento y la consanguinidad. Por lo tanto no se trata de un descubrimiento o de un hallazgo, sino de la aplicación empírica de técnicas genéticas desde hace más de 250 años, que han permitido disponer de animales con un comportamiento característico diferenciado del de los animales bravíos de origen, para lo que se han empleado pruebas funcionales de campo (acoso y derribo y retienta) y observación del comportamiento en plazas de toros.

4.1.2.- Origen histórico

La primera reseña de un festejo taurino aparece en el año 815 en el reino de León. A partir de este momento, los animales destinados a la lidia son elegidos con criterios opuestos a los de la domesticación convencional; se extraen de las “piaras” de bovino los machos más broncos, rebeldes a las formas de manejo de cada época, refractarios a la doma y al yugo, desahuciados para las labores agrícolas, aunque todavía no se reproducen, sino que simplemente se separan del grupo para ser jugados en espectáculos taurinos.

La primera cita de una ganadería de lidia la recogen los historiadores del reino de Navarra y data del siglo XIV: “en 1388 Joan de Gris (ganadero de Tudela) envió 2 toros a Pamplona para fiestas en honor del Duque de Borbón, por mandato del rey Carlos II el Malo”. Según el veterinario D. Justino Pollos

Herrera, “durante los 8 siglos que duró la reconquista, las grandes franjas de territorio que separaban a moros y a cristianos fueron tierras de nadie. Este hecho favoreció al toro silvestre para saltar la Edad media y llegar a nuestros días, incluidos en las grandes posesiones y heredades de los nobles y señores cristianos del recién nacido reino de España”. En gran medida eran toros salvajes cuya población estaba sometida a un cierto control por el hombre, como es el caso de los caballos salvajes gallegos, asturcones y potxokas vascos.

Hasta el siglo XVI, son los carniceros quienes proporcionan toros para la lidia extraídos del vacuno de la tierra. A partir del siglo XVII, y sobre todo durante el siglo XVIII, se produce en España una verdadera especialización productiva en ganaderías de lidia, que no sólo seleccionan este tipo de animales a partir de sus vacadas, sino que además se dedican a la reproducción de los mismos. En la cría del ganado bravo se tienen en cuenta las características individuales de los futuros reproductores, la ascendencia y descendencia para escoger los sementales, y la descendencia para seleccionar vacas. Este sistema de selección, de absoluta vigencia en la actualidad, constituyó en su tiempo una auténtica innovación.

Hasta el siglo XVIII, según asegura el veterinario D. Cesáreo Sanz Egaña (1885-1959), “los toros corridos y muertos en las fiestas procedían de ganaderías vacunas que daban indistintamente animales para el trabajo (bueyes) o reses para la carnicería”.

La corrida de toros surge, según D. José Ortega y Gasset, en torno a 1728 tras un largo proceso evolutivo. Los ganaderos orientan su selección a partir de este momento, y hasta nuestros días hacia este tipo de festejo. Las ganaderías que no se adaptan a las demandas de esta clase de espectáculo, que se encuentra en constante evolución, se van extinguiendo una detrás de otra, a excepción de algunas ganaderías que orientan su selección hacia cierto tipo de festejos taurinos populares.

La de lidia es una de las razas bovinas más antiguas del mundo y pionera en la implantación de un esquema de selección relativamente complejo en el que ya existían registros genealógicos. En esta raza se produce un importante aislamiento reproductivo entre las poblaciones que la constituyen. Hoy en día, podemos observar una gran variabilidad dentro de ella debido al interés de los criadores por disponer de un patrimonio genético propio, que se mantiene durante varias generaciones en el seno familiar. Este objetivo, de alguna manera conservacionista, puede ser contrario al de obtener una mayor rentabilidad económica. El ganadero que adquiere o hereda una ganadería siente la necesidad de conservar y mejorar ese patrimonio genético y de transmitirlo a sus hijos. Gracias a ello se mantienen poblaciones aisladas y se crean encastes, algunos de los cuales están en serio peligro de extinción.

4.2.- Estructura poblacional de la raza de lidia

El toro de lidia tiene una estructura peculiar y única entre las razas de ganado bovino. Por un lado, la fuerte selección que se realiza para caracteres de comportamiento muy definidos y su cría de carácter extensivo en un ecosistema único reducen la intercambiabilidad genética y el cruce con otras razas, por lo que el aislamiento reproductivo está prácticamente garantizado (Cañón *et al.*, 2008).

La existencia desde hace décadas de libros genealógicos, registros exhaustivos de las características morfológicas de cada toro y anotaciones de los distintos intercambios entre ganaderías permite disponer de datos muy precisos sobre la estructura de la raza de lidia.

La raza se divide en distintas líneas, denominadas encastes, cuyos orígenes se remontan al de la propia raza, y los encastes actuales descienden directamente de lo que se conoce como encastes fundacionales. Los encastes engloban generalmente varias ganaderías, aunque se dan casos de encastes con una única ganadería asociada. Además, cada ganadería únicamente cría ejemplares de un determinado encaste, y aunque se producen intercambios de ganado entre distintos encastes buscando un determinado objetivo, no suele ser la práctica más habitual.

Por castas fundacionales se entienden las poblaciones de bovinos de las cuales se extraían toros de lidia, constituidas en ganaderías especializadas en la producción y reproducción de reses de lidia, y que han tenido o podido tener continuidad hasta nuestros días, aunque actualmente puedan aparecer más o menos difusas, como consecuencia fundamentalmente de diversos cruzamientos y otros procesos evolutivos. Las comarcas con mayor número de ganado destinado a la lidia estaban en

Navarra, Castilla y Andalucía, y en menor proporción en Extremadura, Aragón y Portugal, existiendo una relación absoluta entre las zonas de cría y las primitivas castas fundacionales originadas alrededor de las grandes cuencas fluviales, siendo éstas de origen Navarro, Castellano y Andaluz.

Pese a la existencia de muchas más ganaderías en distintos lugares de España de donde se obtenían toros para la lidia, y que incluso se perpetuaron durante varias generaciones de ganaderos, hoy en día se consideran extinguidas y sólo se registran como tales siete castas fundacionales, representadas en recortes de dibujos del pintor Manuel Castellano realizados a mediados del siglo XIX en la Fig. 4.2.1.

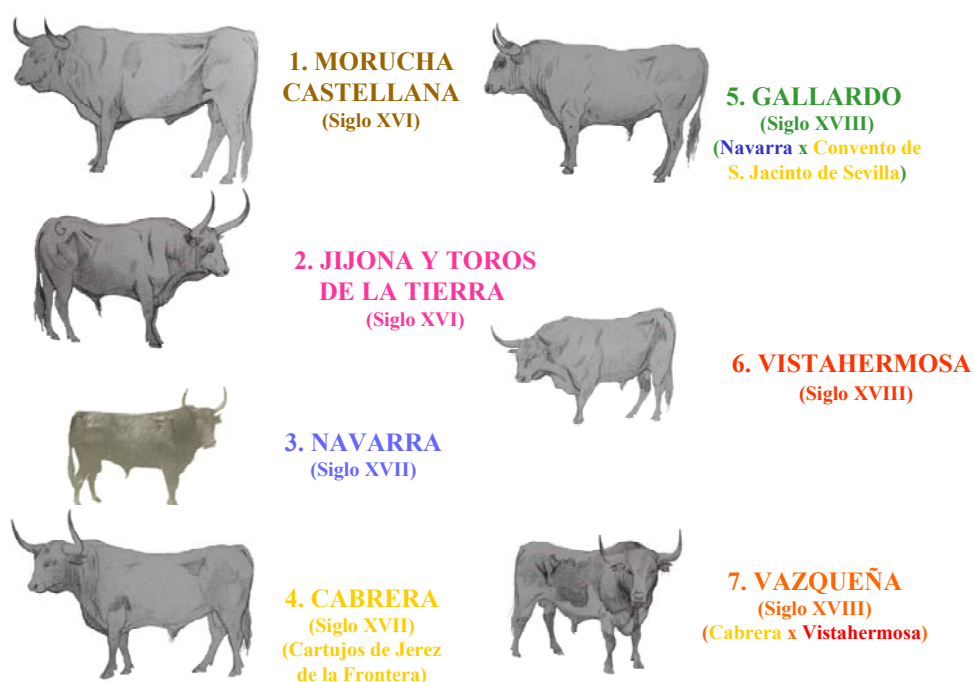


Fig. 4.2.1: Castas fundacionales del toro de lidia

En el origen de las castas andaluzas, Cabrera, Gallardo, Vazqueña y Vistahermosa, influyó considerablemente el ganado frailer (procedente de los diezmos con que los ganaderos pagaban a las congregaciones religiosas), siendo éste de muy diversa procedencia. En el siglo XVIII los conventos y las familias principales realizaban la cría del ganado bravo hasta la llegada de la desamortización, donde el ganado frailer pasa a particulares.

En la casta Navarra existe una marcada influencia del ganado traído por los celtas que se desarrolló en manadas salvajes adquiriendo su condición de bravo. Esta casta constituye una unidad independiente respecto al resto, estando dotada de una idiosincrasia especial.

De la casta Morucha Castellana no existen muchos datos de interés, a pesar de asentarse en una de las zonas con mayor tradición ganadera como es Salamanca.

Finalmente, la casta Jijona (influida por la Real Vacada de Aranjuez en su origen) y Toros de la Tierra (que no constituye una casta propia pero tiene un origen común a la casta Jijona) se sitúa en las dehesas del Jarama al pasar por la comarca madrileña.

Las actuales ganaderías de lidia proceden de estas siete castas fundacionales: Morucha Castellana, Navarra, Jijona y Toros de la Tierra, Cabrera, Vazqueña, Vistahermosa y Gallardo, cuyas diferencias, además de geográficas, también hacen referencia a aspectos morfológicos y de comportamiento.

A partir de estas castas fundacionales es posible establecer un seguimiento de la sucesión de hierros y vacadas, algunas de las cuales se han perdido en pureza. La acción de diferentes fuerzas genéticas y de sistemas de manejo basados en el aislamiento reproductivo entre poblaciones de origen común permite agrupar las ganaderías en encastes. Para que una población se considere un encaste, la(s) ganadería(s) que lo forme(n) debe(n) tener un origen genético conocido y aislado del resto de encastes por un período mínimo de 30 años, además de caracteres de comportamiento o morfológicos propios (Boletín Oficial del Estado, 2001). En el B.O.E. del 13 de febrero del 2001 queda definido el prototipo racial del toro de lidia como el de los diferentes encastes reconocidos. En la figura 4.2.2 se pueden observar ejemplos de morfotipos pertenecientes a distintos encastes actuales, todos ellos parte del estudio realizado en este trabajo.



Carlos Núñez



Murube



Pablo Romero



Torrestrella



Juan Pedro Domecq



Antonio Pérez



Conde de la Corte



Miura



María Montalvo

Fig. 4.2.2: Ejemplos de distintos encastes y sus diferencias morfológicas

5.- Material animal y molecular analizado

En el marco de una serie de estudios destinados a caracterizar genéticamente la raza del toro de lidia español (Silva *et al.*, 2004, 2006; Cortés *et al.*, 2007; Cañón *et al.*, 2007a, 2007b, 2008; Cortés, 2008, Cortés *et al.*, 2008) se tomaron muestras de 1683 animales de 29 encastes con un número variable de ganaderías (desde encastes con una única hasta un máximo de 19) y de muestras (de 11 a 212), pertenecientes todas a la misma generación dentro de cada encaste, y con aproximadamente el mismo número de machos y hembras, y se genotiparon 24 marcadores de tipo microsatélite para cada animal. La recogida de muestras fue realizada por la Unión de Criadores del Toro de Lidia (U.C.T.L.) a través de sus veterinarios, y el genotipado se llevó a cabo en el Laboratorio de Genética de la Facultad de Veterinaria de la U.C.M. Detalles técnicos sobre el proceso de genotipado se pueden encontrar en Cañón *et al.* (2008) y Cortés (2008).

La Tabla 5.1 muestra el número de ganaderías, el número de animales muestreados y las abreviaturas de cada uno de los encastes seleccionados.

Encaste (acrónimo)	Nº de ganaderías	Nº de muestras	Encaste (acrónimo)	Nº de ganaderías	Nº de muestras
Antonio Pérez (ANT)	1	45	Juan Pedro Domecq (DOM)	9	212
Arauz de Robles (ARA)	1	52	Manuel Arranz (MAN)	1	32
Atanasio Fernández (ATA)	6	97	María Montalvo (MON)	1	11
Baltasar Iban (BAL)	2	52	Marqués de Albasarreda (ALB)	3	46
Braganza (BRA)	1	25	Marqués de Villamarta (VIL)	3	60
Carlos Núñez (CAR)	6	71	Miura (MIU)	1	46
Concha y Sierra (SIE)	1	49	Murube (MUR)	4	51
Conde de la Corte (COR)	1	30	Pablo Romero (PAB)	1	50
Conde Santa Coloma (COL)	10	182	Pedrajas (PED)	2	48
Contreras (CON)	4	59	Saltillo (SAL)	4	54
Cuadri (CUA)	1	50	Torrestrella (TOR)	1	50
Félix Gómez (FEL)	1	46	Urcola (URC)	1	23
Gamero Cívico (GAM)	4	57	Vega Villar (VEG)	3	46
Hidalgo Barquero (HID)	4	57	Veragua (VER)	1	32
José Marzal (CRM)	1	50	Total	79	1683

Tabla 5.1: Encastes, número de ganaderías y tamaños muestrales utilizados

En la Tabla 5.2 se detallan los marcadores utilizados, así como el número de alelos detectados y su localización cromosómica. El número de alelos varía entre 6 y 14, y los marcadores están repartidos por todo el genoma, con 22 de los 30 cromosomas representados.

Microsatélite	Nº de alelos	Cromosoma	Microsatélite	Nº de alelos	Cromosoma
bm 1824	6	1	cssm 66	11	14
bm 2113	9	2	hel 1	6	15
inra 23	9	3	inra 35	7	16
rm 188	11	4	tgla 53	15	16
eth 10	7	5	eth 185	8	17
bm 143	10	6	tgla 227	11	18
ilsts 006 9		7 eth	3	11	19
hel 9	10	8	tgla 126	6	20
eth 225	6	9	tgla 122	14	21
inra 37	8	10	hel 5	9	21
bms 2057	10	12	haut 24	9	22
agla 232	13 13		drb	19 23	

Tabla 5.2: Microsatélites utilizados, número de alelos y localización cromosómica

6.- Análisis, resultados y discusión

Se detectaron un total de 234 alelos entre los 24 loci y 1683 individuos genotipados en los 29 encastes. La Tabla 6.1 muestra los valores obtenidos para diversos indicadores de variabilidad dentro de encastes: heterocigosis esperada, heterocigosis observada, número medio de alelos por marcador promediado en todos los marcadores, riqueza alélica y F_{IS} . La heterocigosis esperada varió entre 0.46 (Cuadri) y 0.68 (Contreras). Conde de la Corte, con la segunda heterocigosis más baja, también presentó el menor número de alelos y riqueza alélica, en contraposición a Conde de Santa Coloma, con el mayor número de alelos, y Contreras, con la mayor riqueza alélica. Las correlaciones entre la heterocigosis y el número de alelos por un lado, y la riqueza alélica por otro, fueron de 0.73 y 0.86, respectivamente.

Las discrepancias entre la heterocigosis esperada y observada se miden mediante el F_{IS} . Valores negativos de este parámetro indicarían mayores niveles de heterocigosis observada que los esperados bajo equilibrio, lo que puede ser consecuencia del denominado efecto Wahlund (Hartl y Clark, 1989) si los animales analizados de ese encaste son resultado de cruzamientos recientes entre animales pertenecientes a líneas o ganaderías genéticamente distantes del mismo encaste, como puede haber ocurrido en Urcola o Conde de la Corte, o provenientes de otros encastes, como puede ser el caso de Manuel Arránz o María Montalvo, que han cruzado ganado de origen Jijona con otro proveniente del origen Vistahermosa. En algunos casos estos excesos de heterocigosis podrían ser el resultado de apareamiento asociativo negativo que es posible en algunas ganaderías en las que se ha tratado de compensar caracteres no deseados. En algunos encastes se observan valores de F_{IS} relativamente elevados (> 15%) como es el caso de Conde de Santa Coloma, Juan Pedro Domecq, Gomero Cívico, Hidalgo Barquero, Murube, Pedradas, Saltillo, Vega Villar y Marqués de Villamarta. Estos valores elevados de F_{IS} posiblemente estén indicando una gran diferencia genética entre las ganaderías que integran ese encaste y de las que se han obtenido las muestras para estos análisis. Efectivamente, si en

cada uno de esos encastes se parte la variabilidad genética entre y dentro de ganaderías, se encuentra que de 112 parejas de valores de F_{ST} , sólo una de ellas no es estadísticamente distinta de 0, es decir, casi todo el valor de F_{IS} encontrado para esos encastes es explicado por las diferencias genéticas entre las ganaderías que lo componen.

Encaste	He	Ho	Nº de alelos	Riqueza alélica ¹	F_{IS}
Contreras 0.684		0.589	6.0	5.1	0.140
Conde Santa Coloma	0.673	0.550	7.0	5.0	0.183
Saltillo 0.652		0.507	5.5	4.6	0.223
Concha y Sierra	0.650	0.603	5.1	4.5	0.071*
Carlos Núñez	0.646	0.573	5.8	4.7	0.112
Urcola 0.643		0.593	4.4		0.077*
Veragua 0.638		0.566	5.6	4.1	0.114*
Vega Villar	0.614	0.458	5.0	4.3	0.254
Félix Gómez	0.612	0.583	4.9	4.2	0.047*
Hidalgo Barquero	0.610	0.514	5.2	4.3	0.157
Marqués de Villamarta	0.610	0.517	5.3	4.4	0.153
José Marzal	0.600	0.590	5.3	4.4	0.018*
Miura 0.588		0.525	4.7	4.0	0.107
Braganza 0.587		0.575	4.4		0.02*
Juan Pedro Domecq	0.578	0.489	6.2	4.2	0.153
Baltasar Iban	0.577	0.534	4.7	4.0	0.074
Torrestrella 0.571		0.570	4.8	4.1	0.001*
Pedrajas 0.570		0.485	4.8	3.9	0.148
Pablo Romero	0.569	0.539	4.4	3.8	0.054*
Murube 0.564		0.463	4.8	4.1	0.178
Antonio Pérez	0.560	0.539	4.4	3.8	0.037*
Atanasio Fernández	0.552	0.511	5.0	3.8	0.074
Manuel Arranz	0.548	0.559	4.0	3.7	-0.021*
María Montalvo	0.544	0.548	3.7		-0.008*
Gamero Cívico	0.543	0.433	4.7	3.9	0.203
Arauz Robles	0.536	0.530	4.8	3.9	0.012*
Marqués de Albaserrada	0.528	0.480	4.2	3.4	0.092*
Conde de la Corte	0.465	0.467	3.5	3.1	-0.003*
Cuadri 0.464		0.430	4.3	3.4	0.074*
Promedio entre encastes	0,59	0.53	4.9	4.1	0.115
Total conjunto	0.715	0.524	9.8	5.7	0.267

Tabla 6.1: Parámetros de diversidad dentro de poblaciones. * valores significativamente no distintos de cero. ¹ basada en una muestra mínima de 15 individuos

La Tabla 6.2 refleja los valores de los estadísticos F para los encastes con más de una ganadería.

Encaste	F_{ST}	F_{IS}	F_{IT}
Atanasio Fernández	0.098	0.006	0.103
Baltasar Iban	0.045	0.052	0.095
Carlos Núñez	0.087	0.042	0.126
Conde de Santa Coloma	0.193	0.008	0.200
Contreras 0.174		0.005	0.178
Gamero Cívico	0.194	0.060	0.242
Hidalgo Barquero	0.173	0.024	0.193
Juan Pedro Domecq	0.125	0.048	0.167
Marqués de Albasarreda	0.084	0.036	0.117
Marqués de Villamarta	0.124	0.072	0.188
Murube 0.104		0.109	0.202
Pedrajas 0.194		0.045	0.230
Saltillo 0.259		0.022	0.276
Vega Villar	0.273	0.068	0.322

Tabla 6.2: Estadísticos F por encastes

Por otro lado, el F_{IT} medio, promediado por loci y poblaciones (0.27, IC 95% [0.25-0.30]) resultó bastante elevado, indicando un fuerte déficit global de heterocigotos, lo que se debe por un lado a las desviaciones del equilibrio Hardy-Weinberg indicadas por los altos valores de F_{IS} , y por otro, a la diferenciación entre encastes (F_{ST} promedio entre pares de encastes = 0.18, IC 95% [0.17-0.19]). En este sentido, todos los valores F_{ST} por parejas fueron significativos, variando de 0.05 (Torrestrella y Juan Pedro Domecq) a 0.35 (Marqués de Albaserrada y Conde de la Corte), lo que indica una fuerte diferenciación y aislamiento reproductivo entre los encastes.

Para los análisis de priorización de encastes se utilizó, por un lado, la distancia de Reynolds, escogida por ser la que mejor se ajusta a un modelo como éste, en el que se descarta la influencia de eventos mutacionales en la diferenciación de las poblaciones al estar tratando con tiempos de divergencia cortos. La Tabla 6.3 muestra, para cada encaste, la media de las distancias entre él y el resto, así como el porcentaje de asignación correcta de animales de ese encaste a él mismo utilizando información molecular y siguiendo el procedimiento de Rosberg *et al.* (2001). En promedio, Cuadri aparece como el encaste más aislado, y Contreras, el que menos se diferencia del resto. Como se puede comprobar, la tasa de éxito en la asignación es muy elevada,

con muchos encastes para los que el total de sus animales fueron asignados correctamente a su procedencia sin información previa de otro tipo, lo que indica nuevamente el alto nivel de diferenciación existente entre los encastes. Además, como cabría esperar, el porcentaje de animales bien asignados decrece a medida que lo hace la distancia media de sus respectivos encastes al resto.

Encaste	Distancia promedio	Tasa de asignación correcta	Encaste	Distancia promedio	Tasa de asignación correcta
Cuadri	0.286	100.0	Veragua	0.188	100.0
Marqués Albaserrada	0.282	100.0	Baltasar Iban	0.185	100.0
Pablo Romero	0.268	100.0	Braganza	0.182	96.0
Miura 0.261		100.0	José Marzal	0.179	98.0
Arauz Robles	0.243	100.0	Urcola	0.179	89.5
Conde Corte	0.228	100.0	Conde Santa Coloma	0.172	96.5
Vega Villar	0.220	97.8	Torrestrella	0.172	90.0
Manuel Arranz	0.219	100.0	Marqués Villamarta	0.167	90.0
Félix Gómez	0.206	97.8	María Montalvo	0.166	63.6
Gamero Cívico	0.200	100.0	Saltillo	0.166	100.0
Murube	0.200	100.0	Juan Pedro Domecq	0.159	87.7
Antonio Pérez	0.192	95.6	Hidalgo Barquero	0.151	91.2
Concha y Sierra	0.192	93.9	Carlos Núñez	0.137	94.9
Pedrajas 0.192		100.0	Contreras	0.134	93.3
Atanasio Fernández	0.188	87.6			

Tabla 6.3: Distancia de Reynolds promedio de cada encaste con el resto, y proporción de animales de cada encaste asignados correctamente a él.

El algoritmo de agrupación utilizado para calcular la diversidad de Weitzman proporciona la conversión de la matriz de distancias de Reynolds en otra con propiedades ultramétricas, de forma que la agrupación resultante es la que mejor explicaría el proceso de divergencia que ha producido la actual situación en términos de diversidad (Weitzman, 1992). El dendograma obtenido así (Fig. 6.1) muestra a los encastes agrupados en dos grandes grupos, en los que la longitud de las ramas corresponde al nivel de endogamia, y según el cual, encastes como Cuadri, Arauz de Robles,

Marqués de Albaserrada o Pablo Romero estarían más separados por ser más endogámicos, y por tanto aportarían más diversidad entre encastes.

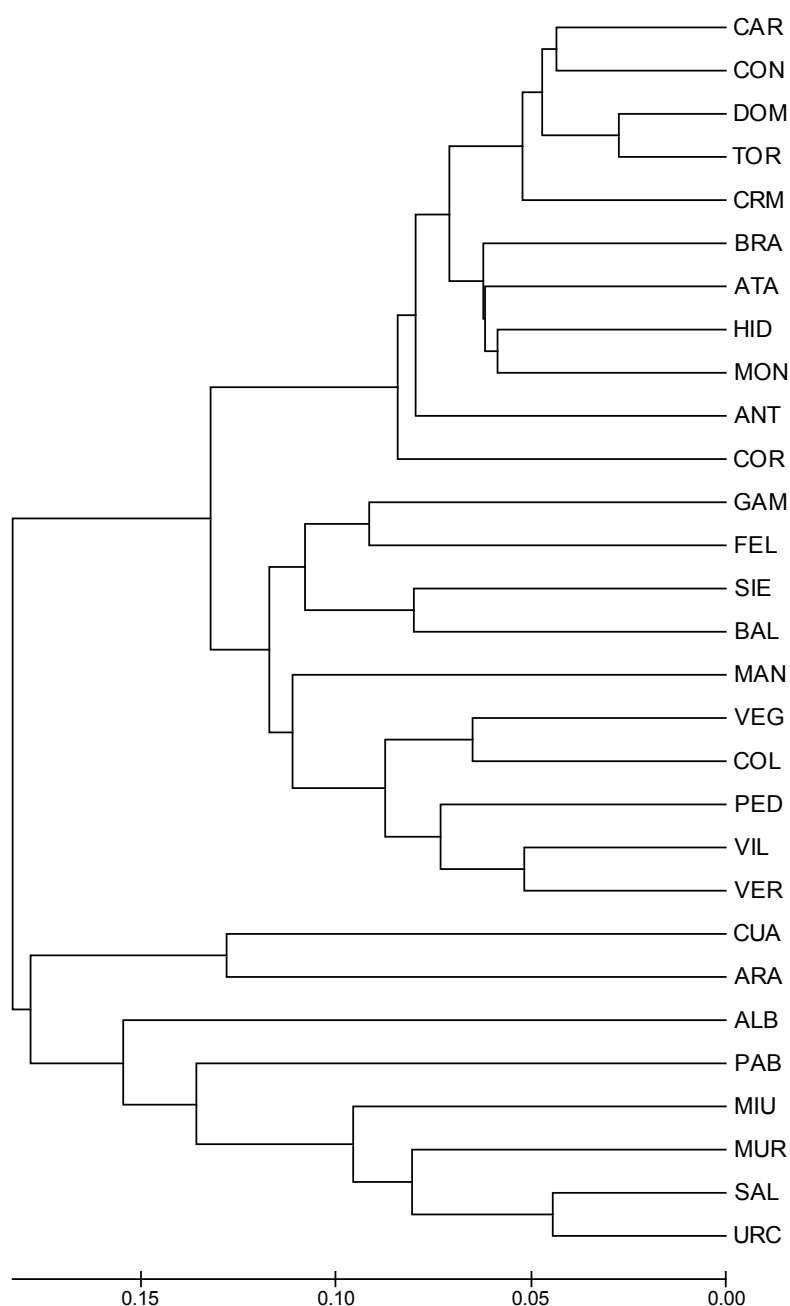


Fig 6.1: Dendrograma realizado con las distancias de Reynolds convertidas a ultramétricas mediante el algoritmo de Weitzman

La otra fuente de información utilizada para el establecimiento de prioridades fueron los parentescos medios estimados mediante marcadores dentro y entre encastes, calculados a partir de similitudes obtenidas de las

frecuencias alélicas (Eding y Meuwissen, 2001). La Tabla 6.4 muestra los parentescos medios dentro de cada encaste y el promedio de los parentescos medios entre cada encaste y el resto.

Encaste	Parentesco medio dentro	Parentesco medio entre	Encaste	Parentesco medio dentro	Parentesco medio entre
Cuadri 0.542		0.280	Miura	0.420	0.226
Conde de la Corte	0.537	0.315	Juan Pedro Domecq	0.419	0.313
Maria Montalvo	0.479	0.321	Veragua	0.412	0.276
Marqués Albaserrada	0.477	0.246	José Marzal	0.408	0.285
Arauz Robles	0.469	0.267	Urcola	0.400	0.278
Manuel Arranz	0.462	0.283	Félix Gómez	0.396	0.257
Gamero Cívico	0.456	0.296	Hidalgo Barquero	0.396	0.300
Braganza	0.452	0.302	Marqués de Villamarta	0.395	0.285
Atanasio Fernández	0.449	0.303	Vega Villar	0.388	0.247
Antonio Pérez	0.448	0.296	Carlos Núñez	0.360	0.290
Murube 0.444		0.290	Saltillo	0.357	0.268
Torrestrella	0.435	0.304	Concha y Sierra	0.354	0.242
Pedrajas	0.433	0.289	Conde Santa Coloma	0.331	0.242
Pablo Romero	0.432	0.226	Contreras	0.322	0.273
Baltasar Iban	0.429	0.290			

Tabla 6.4: Parentescos moleculares medios por encaste

De nuevo Cuadri se confirma como el encaste más endogámico, además de tener un parentesco relativamente alto con el resto de encastes. Los tres encastes con mayor parentesco medio constan de una única ganadería cada uno, mientras que los dos últimos tienen 10 y 4 ganaderías, respectivamente. Teniendo en cuenta que la mayor parte de la variabilidad dentro de un encaste con varias ganaderías es explicada por las diferencias entre ellas, parece razonable que sean aquellas con menos ganaderías las que mayor parentesco medio tengan. La correlación entre el número de ganaderías de un encaste y su parentesco medio es, como cabría esperar, negativa y media-alta ($r = -0.425$).

La relación entre los distintos encastes se puede apreciar mejor de forma gráfica en un gráfico de contorno como el de la Fig. 6.2, obtenida a

partir de los parentescos moleculares. La reordenación viene dada por la agrupación resultante del árbol NJ obtenido a partir de las distancias de parentesco, y los colores más cálidos indican un mayor parentesco.

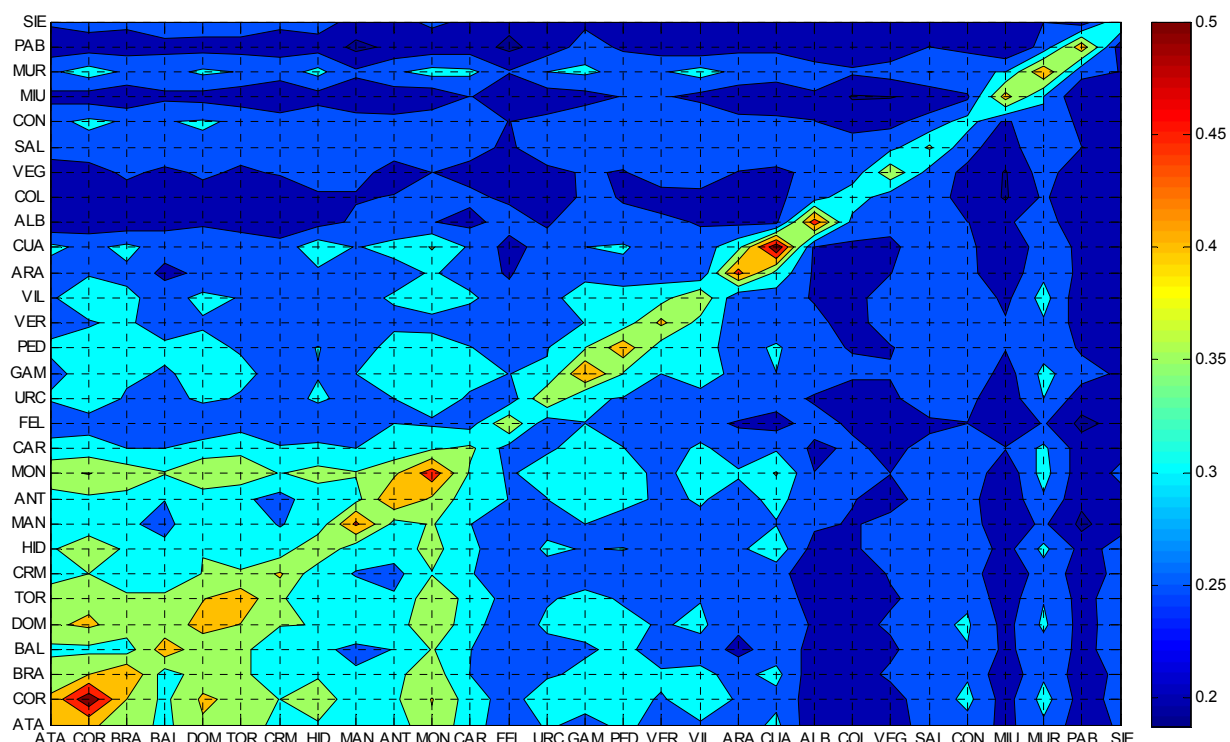


Fig. 6.2: Gráfico de contorno correspondiente a la matriz de parentescos moleculares. Colores más cálidos representan un mayor parentesco

La diagonal aparece perfectamente delimitada, como es lógico, ya que aunque es teóricamente posible que se dé la situación contraria, lo normal es que los parentescos dentro de poblaciones sean mayores que entre ellas. Por otra parte, se ven bien delimitados varios grupos que coinciden parcialmente con lo observado en la figura 6.1, como el de la esquina superior izquierda donde se encuentran encastes como Miura, Murube o Pablo Romero, o más a la izquierda, Arauz de Robles y Cuadri, todos ellos con parentescos moderadamente altos, pero poca relación con otros encastes.

Con el objetivo de comparar las prioridades de conservación obtenidas con los distintos métodos, se calcularon también los parentescos con los métodos de estimación loglineales descritos en la sección 2.4.3, que corrigen

posibles errores en la estimación de probabilidades de identidad en estado en los distintos marcadores y, en general, comportamientos anómalos que se desvíen de las condiciones ideales.

En la Fig. 6.3 presenta el gráfico de contorno correspondiente a la matriz de parentescos obtenida según el método *WLM*, con la ordenación procedente del árbol NJ correspondiente.

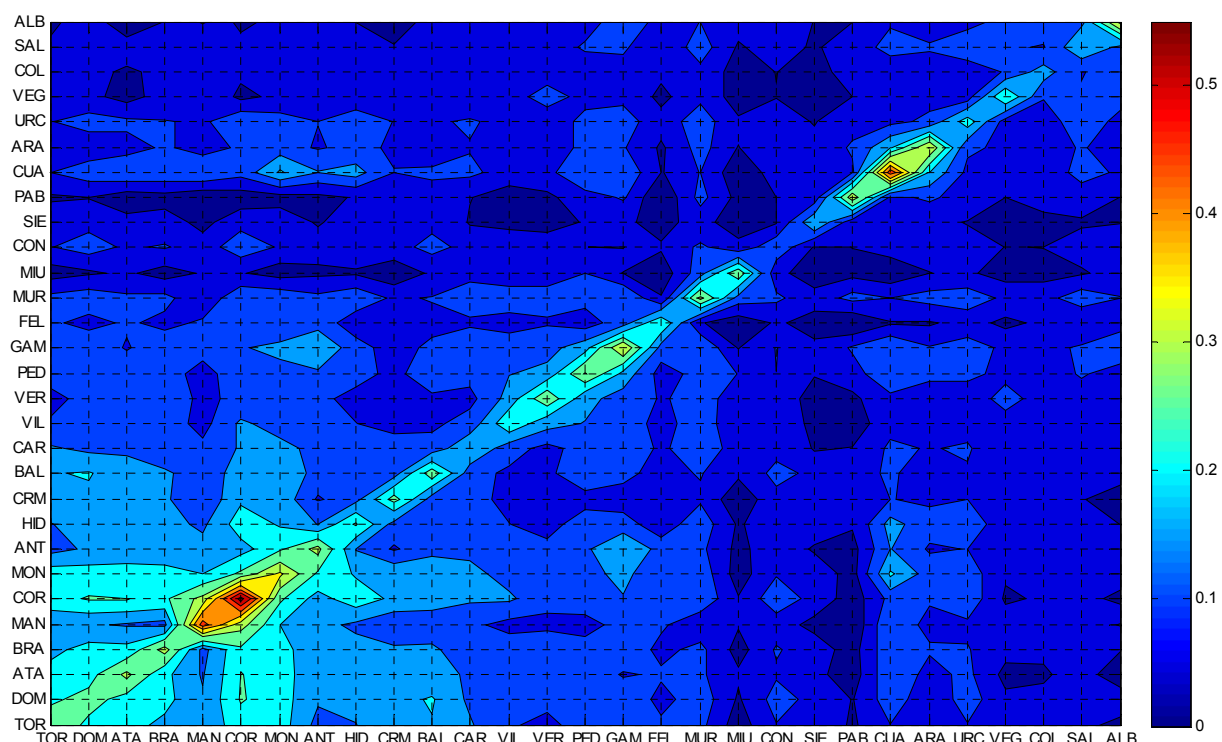


Fig. 6.3: Gráfico de contorno correspondiente a la matriz de parentescos moleculares obtenida con el método WLM.

Se observa una agrupación similar a la anterior, pero con un mayor aislamiento de los grupos entre sí, debido a un menor parentesco entre encastes. Destacan Conde de la Corte y Cuadri como los encastes con mayor parentesco medio dentro de ellos, y el grupo formado por Manuel Arranz, Contreras y María Montalvo, aunque contrariamente a lo que aparece en la Fig. 6.2, hay una mayor relación entre Manuel Arranz y Contreras que entre éste y María Montalvo. Miura aparece de nuevo cerca de Murube, y Cuadri junto a Arauz de Robles.

Más curiosos son los resultados observados con la estimación de los parentescos mediante el método *WLMM*, representados de forma análoga a las dos anteriores en la Fig. 6.4.

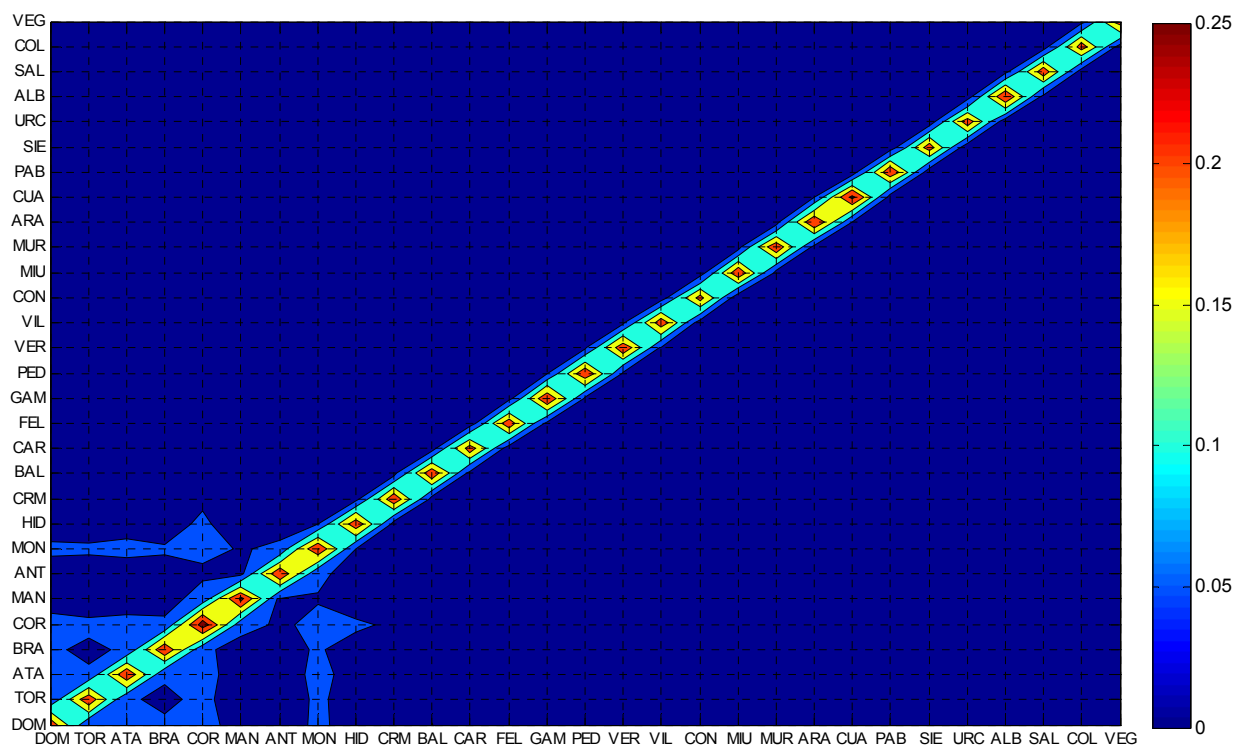


Fig. 6.4: Gráfico de contorno correspondiente a la matriz de parentescos moleculares obtenida con el método *WLMM*.

En este caso el parentesco entre encastes es mucho más reducido. Teniendo en cuenta que el principal objetivo de estimar los parentescos mediante este procedimiento es obtener contribuciones no nulas al core set, se deduce que cuanto mayores sean los parentescos dentro con respecto a los parentescos entre, menos opciones habrá de tener contribuciones negativas en la resolución del sistema lineal del método del core set, y por tanto tener una solución válida en la primera iteración. Podría parecer por lo reducido de los parentescos entre encastes que las contribuciones vendrían dadas en última instancia por las diferencias entre parentescos dentro. Sin embargo, la correlación entre parentescos y contribuciones al core set es, en este caso, alta en valor absoluto, pero no tanto como podría pensarse a la vista de la matriz ($r=-0.57$)

En cuanto a las prioridades de conservación, del método de Weitzman con la distancia de Reynolds se obtienen las contribuciones parciales detalladas en la Tabla 6.5. Las contribuciones con la matriz de distancias original dan mayor peso a aquellos encastes con mayores niveles de endogamia, como era de esperar. Así, Marqués de Albaserrada, Conde de La Corte y Cuadri son los más favorecidos, y a su vez tres de los cuatro encastes más consanguíneos. Las contribuciones parciales calculadas ponderando la matriz de distancias con las heterocigosis esperadas como se indica en el apartado 3.5.3.2 suaviza ligeramente ese efecto. La correlación con la heterocigosis pasa de -0.65 a -0.49 al utilizar este procedimiento.

Encaste	CP _W	CP _{W-He}	Encaste	CP _W	CP _{W-He}
Marques Albaserrada	6.820	5.15	Antonio Perez	2.820	2.86
Conde de La Corte	6.720	4.95	Jose Marzal	2.820	3.26
Cuadri	6.610	5.12	Conde Santa Coloma	2.600	2.97
Pablo Romero	6.160	5.78	Braganza	2.590	2.86
Miura	5.180	5.06	Veragua	2.500	2.66
Arauz de Robles	5.120	4.55	Atanasio Fernandez	2.470	2.5
Manuel Arranz	4.910	4.41	Hidalgo Barquero	2.380	2.45
Vega Villar	3.990	3.64	Maria Montalvo	2.350	2.31
Murube	3.670	3.45	Marques de Villamarta	1.850	2.06
Felix Gomez	3.660	3.75	Carlos Nuñez	1.800	2.03
Baltasar Iban	3.480	3.36	Saltillo	1.770	2.03
Gamero Civico	3.390	3.11	Contreras	1.740	2.16
Concha y Sierra	3.200	3.87	Torrestrella	1.680	1.74
Pedrajas	2.920	3.08	Juan Pedro Domecq	1.100	1.12
Urcola	2.870	3.22			

Tabla 6.5: Contribuciones parciales a la diversidad de Weitzman de los distintos encastes. CP_W: calculadas con la matriz de distancia de Reynolds. CP_{W-He}: idem, matriz ponderada por las heterocigosis esperadas. Se muestra en orden descendiente por CP_W

Aunque el número de encastes (29) es suficientemente bajo como para poder calcular la diversidad de Weitzman y las contribuciones de cada encaste de forma exacta, se realizaron también cálculos siguiendo la aproximación propuesta en García *et al.* (2005) para distintos valores del umbral del coeficiente de variación de la matriz de distancias. La Fig. 6.5 muestra el porcentaje de error en la estimación de la diversidad ([valor

estimado – valor real] / valor real) para seis valores distintos de dicho umbral, con las distancias ponderadas por las heterocigosis medias de cada par.

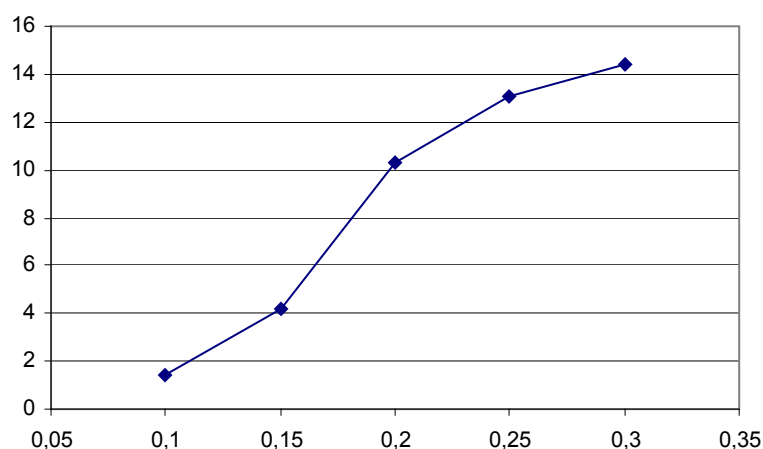


Fig. 6.5: Porcentaje de error en la estimación de la diversidad de Weitzman para distintos valores del umbral del coeficiente de variación en el algoritmo de aproximación

Como se puede apreciar, el error cometido con un umbral que lance la aproximación para un coeficiente de variación inferior a 0.1 es inferior al 2%, mientras que con la aproximación más burda, con un umbral de 0.3, se incurriría en un error de casi el 15%. Mientras que el algoritmo exacto consume aproximadamente 1h 24', la aproximación con el umbral en 0.1 tarda cinco minutos y medio en completarse, lo que supone una ganancia más que importante con una pérdida de precisión reducida. Si se asume un error del 4%, valor aún razonable, con un umbral de 0.15, el tiempo de computación se reduce aún más drásticamente a poco más de 9 segundos.

Si se calculan las correlaciones entre las contribuciones obtenidas utilizando el algoritmo aproximado y las reales, obtenidas con el algoritmo exacto, se puede comprobar que en el peor de los casos, con un umbral de 0.3, la correlación es ya superior a 0.9 ($r=0.92$), mientras que con el umbral situado en 0.1 la correlación supera el 0.99, lo que demuestra el buen comportamiento de la aproximación.

Siguiendo la metodología de Caballero y Toro (2002) (ver también Fabuel *et al.*, 2004), se pueden calcular las proporciones en las que cada

encaste contribuye al valor total de la diversidad actual, entendida ésta como el opuesto al parentesco medio. Para ello se calculó para cada encaste la distancia mínima de Nei media respecto al resto de encastes, y se utilizó junto con el parentesco molecular medio dentro de cada encaste y los tamaños muestrales correspondientes. De todo ello da cuenta la Tabla 6.6.

Encaste	Ni	Parentesco dentro	DNeiProm	Cont. a f	Cont. a D	Cont. Prop.	Cont Prop*
Conde Santa Coloma	182	0.331	0.133	0.021	0.087	12.01	3.86
Juan Pedro Domecq	212	0.419	0.105	0.040	0.086	11.96	3.30
Atanasio Fernandez	97	0.449	0.128	0.019	0.039	5.42	3.27
Carlos Nuñez	71	0.36	0.1	0.011	0.031	4.32	3.56
Contreras 59		0.322	0.098	0.008	0.027	3.77	3.74
Marques de Villamarta	60	0.395	0.12	0.010	0.026	3.58	3.49
Saltillo 54		0.357	0.12	0.008	0.024	3.39	3.67
Hidalgo Barquero	57	0.396	0.107	0.010	0.024	3.33	3.42
Gamero Civico	57	0.456	0.139	0.011	0.023	3.20	3.29
Concha y Sierra	49	0.354	0.143	0.006	0.023	3.18	3.80
Pablo Romero	50	0.432	0.195	0.007	0.023	3.14	3.67
Arauz de Robles	52	0.469	0.172	0.009	0.022	3.01	3.39
Baltasar Iban	52	0.429	0.131	0.009	0.022	3.00	3.38
Jose Marzal	50	0.408	0.127	0.008	0.021	2.96	3.46
Murube 51		0.444	0.139	0.009	0.021	2.92	3.35
Miura 46		0.42	0.189	0.006	0.021	2.91	3.70
Vega Villar	46	0.388	0.154	0.006	0.021	2.90	3.69
Felix Gomez	46	0.396	0.148	0.007	0.021	2.85	3.62
Torrestrella 50		0.435	0.121	0.009	0.020	2.82	3.30
Pedrajas 48		0.433	0.135	0.008	0.020	2.77	3.38
Marqués Albaserrada	46	0.477	0.197	0.008	0.020	2.72	3.47
Cuadri 50		0.542	0.194	0.010	0.019	2.68	3.14
Antonio Perez	45	0.448	0.135	0.008	0.018	2.54	3.31
Veragua 32		0.412	0.138	0.005	0.014	1.91	3.50
Manuel Arranz	32	0.462	0.154	0.006	0.013	1.82	3.33
Conde de La Corte	30	0.537	0.158	0.007	0.011	1.53	2.99
Braganza 25		0.452	0.131	0.005	0.010	1.40	3.27
Urcola 23		0.4	0.129	0.004	0.010	1.38	3.51
Maria Montalvo	11	0.479	0.125	0.002	0.004	0.58	3.11

Tabla 6.6: Para cada encaste, tamaño de muestra, parentesco medio dentro de encaste, promedio de la distancia mínima de Nei con el resto de encastes, contribución absoluta al parentesco global, calculada como $N_i/N_{TOT} \cdot (\text{Parentesco dentro} - D_{\text{neiProm}})$, contribución absoluta a la diversidad, calculada como N_i/N_{TOT} menos el valor en la columna anterior, contribución proporcional a la diversidad y contribución proporcional ignorando el tamaño muestral (Fabuel *et al.*, 2004)

Asumir las contribuciones proporcionales de la penúltima columna como válidas equivale a asumir que la proporción en la que cada encaste está presente en la muestra es la misma en la que está presente en la población completa. De no ser así, han de hacerse correcciones, aunque existe cierta controversia al respecto (Fabuel *et al.*, 2004). Lo que sí es claro es que la proporción de diversidad atribuible a cada subpoblación es muy dependiente

de la representación censal que ésta tenga con respecto al total. Ignorar los tamaños muestrales, como en Fabuel *et al.* (2004) lleva a contribuciones mucho más homogéneas (ver última columna de la tabla anterior) que en la práctica no resultarían demasiado útiles para distribuir recursos.

Seguidamente se aplicó la metodología de Eding *et al.* (2002), utilizando tres matrices de parentescos distintas. En primer lugar, la obtenida directamente a través de frecuencias alélicas y ya detallada anteriormente. A continuación se calculó la correspondiente solución al modelo *WLM*, en el que se tienen en cuenta las diferencias entre los distintos loci, y se procedió análogamente para resolver el modelo *WLMM*. Los resultados correspondientes se muestran en la Tabla 6.7.

Encaste	CP _{CS}	% P _{CS}	CP _{CSWLM}	% P _{CSWLM}	CP _{CSWLMM}	% P _{CSWLMM}
Marques Albaserrada	0	5.285	0	4.873	4.238	0.044
Antonio Perez	0	4.902	0	4.553	2.435	0.013
Arauz Robles	1.239	5.634	0	4.865	3.386	0.027
Atanasio Fernandez	0	4.093	0	4.651	2.406	0.012
Baltasar Iban	0	5.167	0	4.684	2.319	0.012
Braganza 0		3.996	0	4.541	2.126	0.01
Carlos Nuñez	0	4.968	0	4.769	3.055	0.02
Conde Santa Coloma	19.82	7.153	11.245	6.203	5.976	0.081
Contreras 1.6		4.901	14.139	4.521	5.232	0.06
Conde de La Corte	0	4.901	0	4.521	0.056	0
Jose Marzal	0.355	4.901	2.464	4.521	3.937	0.034
Cuadri	2.637	4.901	1.57	4.521	2.483	0.015
Juan Pedro Domecq	0	4.901	0	4.521	0.999	0.002
Felix Gomez	12.041	4.901	13.612	4.521	6.255	0.092
Gamero Civico	0	4.901	0	4.521	0.858	0.002
Hidalgo Barquero	0	4.901	0	4.521	2.343	0.012
Manuel Arranz	0.646	4.901	0.518	4.521	3.208	0.025
Miura	17.193	4.901	13.769	4.521	7.979	0.159
Maria Montalvo	0	4.901	0	4.521	0	0
Murube	0	4.901	0	4.521	1.091	0.003
Pablo Romero	13.325	4.901	7.734	4.521	6.772	0.16
Pedrajas	0	4.968	0	4.521	1.598	0.006
Saltillo 0		4.901	1.363	4.521	4.279	0.04
Concha y Sierra	19.194	4.968	18.748	4.769	7.758	0.144
Torrestrella	0	4.901	0	4.521	2.285	0.011
Urcola	4.074	4.901	0	4.521	3.528	0.028
Vega Villar	7.875	4.901	9.826	4.521	5.85	0.129
Veragua	0	4.901	0	4.521	3.088	0.021
Marques de Villamarta	0	4.968	5.012	4.521	4.462	0.044

Tabla 6.7: Contribuciones y pérdidas relativas de diversidad para el método del Core Set con tres matrices de parentescos moleculares: la simple (CP_{CS}, %P_{CS}), la obtenida con el modelo loglineal *WLM* (CP_{CSWLM}, %P_{CSWLM}), y la obtenida con el modelo loglineal mixto *WLMM* (CP_{CSWLMM}, %P_{CSWLMM}).

En la tabla se pueden observar un buen número de contribuciones nulas al core set, tanto en el procedimiento con las estimaciones estándar de los parentescos como en el que utiliza la matriz estimada por el método *WLM*. Esto es producto de las sucesivas iteraciones realizadas para eliminar las contribuciones negativas que aparecen en la resolución analítica de las ecuaciones lineales. Consecuentemente, la diversidad conservada en el core set con las contribuciones finales es menor que la que se tendría con las contribuciones óptimas. En la Fig. 6.6 se puede observar esta reducción en diversidad a medida que avanzan las iteraciones.

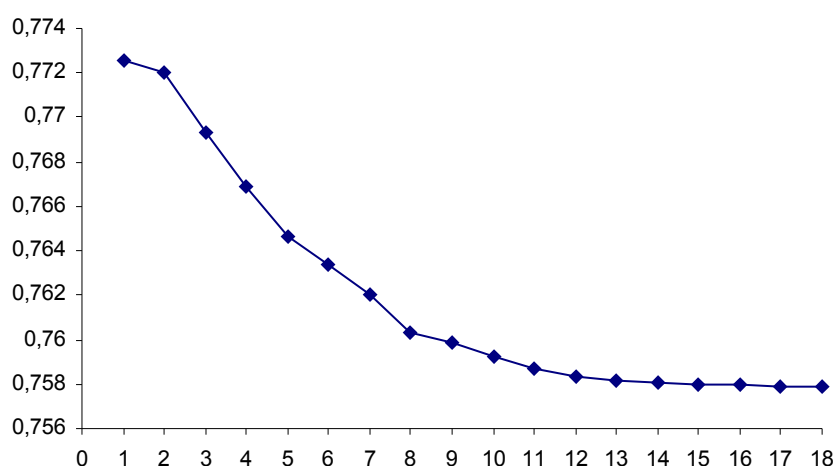


Fig. 6.6: Valor de la diversidad conservada en el core set (ordenadas) en cada iteración del algoritmo (abscisas)

Además de las contribuciones al conjunto nuclear, en la tabla se muestran las pérdidas de diversidad que generarían las eliminaciones de cada uno de los encastes. El modelo *WLMM* se puede ajustar para tener un número máximo determinado de contribuciones nulas a través del parámetro de regresión λ , y en este caso se optó por resolverlo de forma que como mucho haya una única contribución nula y el resto sean positivas. La utilización de este procedimiento proporciona una mejor forma de priorizar las diferentes unidades, ya que es habitual que con la metodología original un número alto de contribuciones sean nulas, imposibilitando así la forma de distinguir entre las unidades correspondientes.

Ambos modelos loglineales corrigen posibles sesgos debidos a malas estimaciones de los parentescos y a comportamientos irregulares de los marcadores. Esto explicaría contribuciones tan dispares como las de Contreras, Urcola, Pablo Romero o Marqués de Villamarta. Otros encastes, como Conde de Santa Coloma, Félix Gómez, Miura o Concha y Sierra, mantienen unos importantes niveles de contribución al core set, debido fundamentalmente a los bajos niveles de parentesco que presentan. El *WLMM* permite además mitigar el efecto que el parentesco dentro de poblaciones tiene en las contribuciones, permitiendo a encastes como Marqués de Albaserrada entrar a contribuir en una proporción significativa. Un estudio más pormenorizado requeriría examinar cada una de las matrices de parentescos en cada uno de los marcadores hasta detectar la causa de las alteraciones.

He							
CoreNormal	0,388						
CoreWLM	0,540	0,842					
CoreWLMM	0,529	0,806	0,848				
CPWeitz	-0,655	0,207	0,044	0,145			
CPWeitzHe	-0,492	0,390	0,213	0,336	0,960		
Contribución proporcional	0,281	0,283	0,173	0,109	-0,325	-0,342	
Cont prop ignorando tamaño muestral	0,812	0,693	0,753	0,879	-0,175	0,014	0,294

Tabla 6.8: Correlaciones entre las contribuciones a la diversidad obtenidas según los métodos denotados en la primera columna.

La tabla 6.8 muestra las correlaciones entre las contribuciones obtenidas mediante estos métodos, observándose que todos ellos, a excepción del de Weitzman, correlacionan positivamente con la heterocigosis, siendo el de las contribuciones proporcionales ignorando el tamaño muestral el que mayor correlación presenta con ella ($r = 0.812$). Las correlaciones negativas de las contribuciones de Weitzman eran de esperar, dada la influencia de la diversidad entre unidades en este método. Sin embargo, la corrección propuesta consistente en ponderar las distancias por las heterocigosis medias entre pares de unidades hace disminuir la negatividad de la correlación de las contribuciones sin ponderar, tanto con la heterocigosis como, lógicamente, con el resto de contribuciones que correlacionan positivamente con ella. Las sucesivas modificaciones en la forma de estimar

parentescos provocan asimismo un aumento en la correlación con la heterocigosis al ser aplicadas en el método del core set.

Finalmente se calcularon las contribuciones de cada encaste a la heterocigosis media, eliminando sucesivamente cada uno de los encastes y comparando la heterocigosis media con la obtenida para el conjunto total, de acuerdo con lo expuesto en el apartado 3.3, de forma que se obtiene la pérdida relativa de heterocigosis al eliminar cada encaste. Se combinaron entonces las contribuciones según la diversidad de Weitzman, en las que sólo interviene la variabilidad entre encastes, con las contribuciones a la heterocigosis, que sólo tienen en cuenta la variabilidad dentro de encastes, ponderando en primer lugar por el F_{ST} (0.18) y a continuación otorgando 5 veces más peso a las contribuciones entre que a las contribuciones dentro. Los resultados se detallan en la Tabla 6.9.

Encaste	CP _{He} (%)	CP _{We}	CP _{Fst}	CP _{5/1}
Conde Santa Coloma	2.127	0.026	0.022	0.025
Marques Albaserrada	0.104	0.068	0.013	0.057
Miura	0.186	0.052	0.011	0.043
Cuadri	-0.172	0.066	0.010	0.055
Vega Villar	0.392	0.040	0.010	0.034
Arauz de Robles	0.056	0.051	0.010	0.043
Conde de La Corte	-0.339	0.067	0.009	0.055
Manuel Arranz	-0.140	0.049	0.008	0.041
Felix Gomez	0.128	0.037	0.008	0.031
Juan Pedro Domecq	0.604	0.011	0.007	0.010
Murube	0.022	0.037	0.007	0.031
Concha y Sierra	0.084	0.032	0.006	0.027
Baltasar Iban	-0.103	0.035	0.005	0.029
Urcola	0.024	0.029	0.005	0.024
Contreras	0.265	0.017	0.005	0.015
Pedrajas	-0.005	0.029	0.005	0.024
Carlos Nuñez	0.240	0.018	0.005	0.015
Gamero Civico	-0.121	0.034	0.005	0.028
Jose Marzal	-0.026	0.028	0.005	0.023
Saltillo	0.193	0.018	0.005	0.015
Veragua	-0.004	0.025	0.004	0.021
Antonio Perez	-0.122	0.028	0.004	0.023
Marques de Villamarta	0.072	0.019	0.004	0.016
Braganza	-0.120	0.026	0.004	0.021
Hidalgo Barquero	-0.100	0.024	0.003	0.020
Maria Montalvo	-0.102	0.024	0.003	0.019
Atanasio Fernandez	-0.319	0.025	0.002	0.020
Torrestrella	-0.172	0.017	0.002	0.014
Pablo Romero	-1.511	0.062	-0.001	0.049

Tabla 6.9: Contribuciones a la heterocigosis media (en %), contribuciones a la diversidad de Weitzman y contribuciones combinadas, con ponderación igual al F_{ST} (ordenadas según esta contribución) y a cinco veces más a la diversidad entre.

Se observan contribuciones negativas a la heterocigosis, provenientes del hecho de estar utilizando valores promediados, y del solapamiento genético existente entre los distintos encastes. Algunos con valores muy negativos presentan a su vez valores de heterocigosis dentro de ellos muy bajos en comparación con el resto, como Cuadri o Conde de la Corte. Hay otros encastes, sin embargo, cuya desaparición también conllevaría un aumento de la heterocigosis media, como Pablo Romero, Torrestrella o Hidalgo Barquero, que tienen valores de heterocigosis interna medios o altos.

Lo que se observa a la vista de las contribuciones combinadas ponderadas por el F_{ST} es que bien podrían interpretarse como un reajuste de las contribuciones de Weitzman para tener en cuenta a los encastes con mayor contribución a la heterocigosis media. Llamen la atención en este sentido las diferencias de Conde de Santa Coloma, con la mayor contribución combinada, pero en un ranking en la parte media-baja de la tabla de contribuciones de Weitzman, y de Juan Pedro Domecq, el que menos aporta a la diversidad de Weitzman, y sin embargo con una contribución combinada media-alta. La ordenación resultante de la combinación con la ponderación 5/1 es prácticamente idéntica a la producida utilizando únicamente la diversidad de Weitzman. De hecho, la correlación entre estas dos contribuciones es superior a 0.99, mientras que la de Weitzman con la combinada ponderada por el F_{ST} es de 0.33, y la de ésta con la obtenida con la heterocigosis es de 0.77. En cualquier caso, no se puede decir que haya una forma universalmente mejor de ponderar, y dependerá de la intención con la que se aborde el programa de conservación el elegir unos pesos u otros. La gran ventaja de esta forma de obtener contribuciones radica justamente en la posibilidad de precisar la importancia relativa que se le da a cada uno de los dos componentes de la diversidad, y esta arbitrariedad, por tanto, puede ser vista como una ventaja o como un inconveniente.

Para una discusión detallada sobre las medidas tradicionales de diversidad y diferenciación entre encastes (heterocigosis, riqueza alélica y estadísticos F) se recomienda consultar Cañón *et al.* (2008). En cuanto a la

asignación de prioridades, lo cierto es que las particularidades de la raza de lidia emergen una vez más cuando se trata de comparar unos encastes con otros en términos de contribución a la diversidad. Parece que hay una tendencia común a favorecer encastes de una sola ganadería, como Concha y Sierra, Miura o Pablo Romero, que pese a seguir prácticas consanguíneas destinadas a fijar caracteres, y alejarse genéticamente del resto de encastes, mantienen niveles aceptables de heterocigosis. Esta tendencia es más difícil de observar en las contribuciones de Caballero y Toro (2002) ponderadas por los tamaños muestrales, pero vuelven a aparecer cuando se asumen tamaños iguales.

Aún así este procedimiento debería ser utilizado con cautela y disponiendo de la mejor información posible sobre la relación entre tamaños muestrales y poblacionales. Se pueden encontrar discusiones sobre correcciones por tamaños de muestra en Nei y Chessser (1983) Pons y Chaouche (1995) y Excoffier (2001).

Por otra parte, encastes como Antonio Pérez, Braganza, José Marzal, María Montalvo o Torrestrella se ven también uniformemente perjudicados en el establecimiento de prioridades. Todos ellos tienen en común el ser encastes con varias ganaderías dentro de ellos muy diferenciadas entre sí y con altos niveles de endogamia debido al poco intercambio de material genético entre ellas.

Futuras investigaciones deberían ir encaminadas a la actuación dentro de encastes en estas condiciones, analizar la situación genética con procedimientos de priorización de este tipo, y quizás poder llegar a plantearse una política de cruzamientos entre ganaderías para generar heterocigosis.

PARTE III

Conclusiones

No es esta una tesis que, por concepto, se preste a unas conclusiones puntuales. El trabajo que hay detrás de ella no va encaminado a la consecución de unos objetivos singularmente determinados, sino que pretende ser una herramienta de consulta y punto de partida para futuros estudios metodológicos y de aplicación.

Aún así, se ha cumplido con ese objetivo difuso de dar una perspectiva rigurosa de la situación actual de un problema, el de la asignación de prioridades en esquemas de conservación de la diversidad genética, que está generando publicaciones de forma constante en los últimos años, y que sin embargo adolece de una revisión conjunta y en profundidad de los distintos métodos existentes para resolverlo.

Más allá de esa visión conjunta, y sin embargo, profunda, de métodos que habitualmente aparecen aplicados, y sobre todo, estudiados, por separado, se han realizado algunos avances metodológicos, como la inclusión de información intrapoblacional en la diversidad de Weitzman y el desarrollo de una aproximación al algoritmo recursivo original, que aportan soluciones a problemas previos y que han sido oportunamente publicados y utilizados, además de estar sirviendo de base para artículos en curso. Su aplicación en la raza de lidia confirma que sirven bien al propósito con el que fueron desarrollados.

El procedimiento de Weitzman con la matriz de distancias ponderada por las heterocigosis proporciona unos resultados algo más equilibrados que la versión original sin ponderar, y como era de esperar, disminuye la correlación negativa con la heterocigosis, de forma que la diversidad dentro de unidades tiene algo más de peso al calcular las contribuciones. La imposibilidad de disponer de tamaños efectivos en el ejemplo del toro de lidia, por carecer de conexiones descritas entre las distintas ganaderías de los encastes con más de

una, hace que no se hayan podido aplicar las diversidades marginales y elasticidades y estudiar su comportamiento. Habría sido interesante ver cuál habría sido en un futuro a corto, medio y largo plazo.

El cálculo de las marginales y elasticidades no es exclusivo de la diversidad de Weitzman, sino que es aplicable a cualquier otra medida, entre las que se podría encontrar perfectamente la definida a través de los parentescos medios. Sigue pendiente, sin embargo, un estudio más profundo de estas medidas de forma que las diversidades que se calculen para un determinado horizonte temporal tengan en cuenta las evoluciones previstas bajo distintos modelos (contemplando deriva, mutación y/o selección), y no se estén calculando diversidades al cabo de un determinado número de generaciones basadas en distancias o parentescos actuales.

La aproximación del algoritmo permite acometer estudios que de otra forma no serían viables con un coste aparentemente reducido en términos de errores. Su utilización en el ejemplo de la raza de lidia no ha sido necesaria ya que se ha podido aplicar el algoritmo completo al tener un número de unidades intermedio. Sí habría sido necesario de haber podido disponer de la información requerida para calcular diversidades marginales y elasticidades, al tener que estudiar todas las posibles combinaciones contenidas en el conjunto de las partes del total. Aún así, se ha estudiado el comportamiento de la aproximación con distintos niveles de precisión, y se ha comprobado que aún en la situación más desfavorable proporciona resultados bastante similares a los obtenidos con el algoritmo exacto con un consumo computacional netamente inferior.

El enfoque de Caballero y Toro sobre las contribuciones a la diversidad actual ha resultado ser demasiado dependiente de la estructura muestral, y parece que resulta conveniente optar alternativamente por una estrategia de minimización de parentesco para calcular contribuciones óptimas, como ellos mismos plantean y Eding *et al.* llevan a cabo más analíticamente. Los modelos loglineales de estimación de parentescos, y especialmente el mixto ajustado

para minimizar las contribuciones nulas dan resultado, y quizás el *WLMM* resulte más útil en la práctica para establecer órdenes de prioridades.

El ejemplo del toro de lidia no debería verse en esta tesis más que como lo que es: un ejemplo de aplicación de las metodologías que se han descrito. Para un estudio verdaderamente relevante se requieren análisis e interpretaciones más profundos y que cuenten con la ayuda de expertos en la raza para desentrañar los porqués de las contribuciones e ir más allá, realizando análisis dentro de los encastes e incluso dentro de las ganaderías. Pero como no podía ser de otra forma, ya se está trabajando en ello, y futuras publicaciones darán fe de los avances y los resultados obtenidos.

Sin embargo, a pesar del hecho de que la raza de lidia se preste especialmente, por sus peculiaridades únicas, a un estudio intrarraza, la conclusión que debe destilarse de la aplicación de las distintas metodologías en ella es que se abre una vía que quizás debería ser más explotada: la de los estudios de asignación de prioridades dentro de razas. Para ello se pueden considerar como unidades rebaños, ubicaciones geográficas o cualquier otra distinción. Sin ir más lejos, y por seguir con el ejemplo de la raza de lidia, sólo en el encaste Pablo Romero aparecen 15 grupos genéticos claramente determinados cuando se analizan unos 300 animales pertenecientes a él. En un mundo, el de las especies ganaderas, en el que el poner unas razas por delante de otras suscita no poca controversia, y en el que en la práctica se asignan recursos económicos sin atender a este tipo de criterios de priorización, cuya aplicación se queda poco menos que circunscrita a las publicaciones científicas, quizás deberían plantearse estrategias de priorización dentro de razas para que todos los paquetes metodológicos que se desarrollan tengan una aplicación real.

Bibliografía

- Akey, J.M.; Eberle, M.A.; Rieder, M.J.; Carlson, C.S.; Shriver, M.D.; Nickerson, D.A.; Kruglyak, L. (2004) Population history and natural selection shape patterns of genetic variation in 132 Genes. *Plos Biology*, 2: 1591-1599.
- Amos, W.; Balmford, A. (2001) When does conservation genetics matter? *Heredity*, 87: 257-265.
- Anderung, C.; Bouwman, A.; Persson, P.; Carretero, J.M.; Ortega, A.I., Elburg, R.; Smith, C.; Arsuaga, J.L.; Ellegren, H.; Götherström, A. (2005) Prehistoric contacts over the straits of Gibraltar indicated by genetic analysis of Iberian Bronze Age cattle. *Proceedings of the National Academy of Sciences USA*, 24: 8431-8435.
- Aranguren-Méndez, J.; Jordana, J.; Gómez, M. (2002) Genetic conservation of five endangered Spanish donkey breeds. *Journal of Animal Breeding and Genetics*, 119: 256-263.
- Avriel, M. (1976) *Nonlinear Programming, Analysis and Methods*. Prentice Hall, Inc., Englewood Cliffs, N. J.
- Baker, C.M.A., Manwell, D. (1991) Population genetics, molecular markers and gene conservation of bovine breeds. In: Hickman, C.G., ed., *Cattle Genetic Resources*. (1st ed.), pp.221-304. Elsevier, Amsterdam
- Bamshad, M.; Wooding, S.P. (2003) Signatures of natural selection in the human genome. *Nature Reviews Genetics*, 4: 99-111.
- Barker, J.S.F. (1999) Conservation of livestock breed diversity. *Animal Genetic Resources Information*, 25: 33-43.
- Barker, J.S.F. (2001) Conservation and management of genetic diversity: a domestic animal perspective. *Canadian Journal of Forestal Research*, 31: 588-595.
- Barker, J.S.F.; Tan, S.G.; Moore, S.S.; Mukherjee, T.K.; Matheson, J.-L.; Selvaraj, O.S. (2001) Genetic variation within and relationships among populations of Asian goats (*Capra hircus*) *Journal of Animal Breeding and Genetics*, 118: 213-233.
- Baumung, R.; Simianer, H.; Hoffman, I. (2004) Genetic diversity studies in farm animals – a survey. *Journal of Animal Breeding and Genetics*, 121: 361-373.
- Beaumont, M.A.; Balding, D.J. (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular ecology*, 13: 696-980.
- Beja-Pereira, A.; Alexandrino, P.; Bessa, I.; Carretero, Y.; Dunner, S.; Ferrand, N.; Jordana, J.; Laloe, D.; Moazami-Goudarzi, K.; Sanchez, A.; Cañon, J. (2003) Genetic characterization of southwestern European bovine breeds: a historical and

biogeographical reassessment with a set of 16 microsatellites. *Journal of Heredity*, 94:243-250.

Bennewitz, J.; Kantanen, J.; Tapio, I.; Li, M.H.; Kalm, E.; Vilkki, J.; Ammosov, I.; Ivanova, Z.; Kiselyova, T.; Popov, R.; Meuwissen, T.H.E. (2006) Estimation of breed contributions to present and future genetic diversity of 44 North Eurasian cattle breeds using core set diversity measures. *Genetics, Selection and Evolution*, 38: 201-220.

Bennewitz, J.; Meuwissen, T.H.E. (2004) Investigations on the selection of breeds for conservation of genetic variance. 55th Annual Meeting of the European Association for animal Production. Bled, 5 - 8 September 2004.

Bennewitz, J.; Meuwissen, T.H.E. (2005) Estimation of extinction probabilities of five german cattle breeds by population viability analysis. *Journal of Dairy Science*, 88: 2949-2961.

Black, W.C.; Baer, C.F.; Antolin, M.F.; DuTeau, N.M. (2001) Population genomics: genome-wide sampling of insect populations. *Annual Review of Entomology*, 46: 441-469.

Boletín Oficial del Estado (2001) Real Decreto 60/2001, de 26 de enero, sobre prototipo racial de la raza bovina de lidia. *Boletín Oficial del Estado* 38, 5255-61.

Bonin, A.; Nicole, F.; Pompanon, F.; Miaud, C.; Taberlet, P. (2007) Population adaptive index: a new method to help measure intraspecific genetic diversity and prioritize populations for conservation. *Conservation Biology*, 21: 697-708.

Bowen, B.W. (1999) Preserving genes, species, or ecosystems? Healing the fractured foundations of conservation policy. *Molecular Ecology*, 8: S5-S10.

Bradley, D.G.; MacHugh, D.E.; Loftus, R.T.; Sow, R.S.; Hoste, C.H.; Cunningham, E.P. (1994) Zebu-taurine variation in Y chromosomal DNA : a sensitive assay for genetic introgression in west African trypanotolerant cattle populations. *Animal Genetics*, 25: 7-12.

Bradley, D.G.; MacHugh D.E.; Cunningham, P.; Loftus, R.T. (1996) Mitochondrial diversity and the origins of African and European cattle. *Proceedings of the National Academy of Sciences USA*, 93: 5131-5135.

Brinkmann, B.; Klitschar, M.; Neuhuber, F.; Huhne, J.; Rolf, B. (1998) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *American Journal of Human Genetics*, 62: 1408-1415.

Bruford, M.W. (2004) Conservation genetics of UK livestock: from molecules to management. In Simm, G.; Villanueva, B.; Townsend, S., ed.: *Farm Animal Genetic Resources*, pp. 151-169. Nottingham University Press, Nottingham, UK.

Caballero, A. (1994) Developments in the prediction of effective population size. *Heredity*, 73: 657-679.

Caballero, A.; Toro, M.A. (2000) Interrelations between effective population size and other pedigree tools for the management of conserved populations. *Genetical Research Combined*, 75: 331-343.

Caballero, A.; Toro, M.A. (2002) Analysis of genetic diversity for the management of conserved subdivided populations. *Conservation Genetics*, 3: 289-299.

Cañón, J.; Alexandrino, P.; Bessa, I.; Carleos, C.; Carretero, Y.; Dunner, S.; Ferran, N.; Garcia, D.; Jordana, J.; Laloë, D.; Pereira, A.; Sanchez, A.; Moazami-Goudarzi, K. (2001) Genetic diversity measures of local European beef cattle breeds for conservation purposes. *Genetics Selection and Evolution*, 33: 311-332.

Cañón, J.; Cortés, O.; García, D.; García-Atance, M.A.; Tupac-Yupanqui, I.; Dunner, S. (2007a) Distribución de la variabilidad genética en la raza de lidia. *Archivos de Zootecnia*, 56 (Sup. 1): 391-396.

Cañón, J.; Fernández, J.; García-Atance, M.A.; Tupac-Yupanqui, I.; Cortés, O.; García, D.; Dunner, S. (2007b) Relaciones genéticas entre ganaderías de lidia en función de su origen. *Archivos de Zootecnia*, 56 (Sup.1): 383-390.

Cañón, J.; Tupac-Yupanqui, I.; García-Atance, M.A.; Cortés, O.; García, D.; Fernández, J.; Dunner, S. (2008) Genetic variation within the Lidia bovine breed. *Animal Genetics*, doi:10.1111/j.1365-2052.2008.01738.x.

Cavalli-Sforza, L.L.; Edwards, A.W.F. (1967) Phylogenetic analysis: models and estimation procedures. *American Journal of Human Genetics*, 19: 233-257.

Chaiwong, N.; Kinghorn, B.P. (1999) Use of genetic markers to aid conservation decisions for groups of rare domestic breeds. *Proceedings of the Association for the Advancement of Animal Breeding and Genetics*, 13: 365-368.

Cockerham, C.C. (1969) Variance of gene frequencies. *Evolution*, 23: 72-84.

Cooper, L. (1981) *Mathematical Programming for Operations Researchers and Computer Scientists*. A. G. Holtzman, Marcel Dekker, Inc., New York.

Cortés, O.; Dunner, S.; García-Atance, M.A.; García, D.; Cañón, J. (2007) Matrilineas en la formación del ganado bovino de lidia. *Archivos de Zootecnia*, 56 (Sup. 1): 397-402.

Cortés, O. (2008) Análisis de la variabilidad genética en la raza bovina de lidia utilizando información molecular. Tesis Doctoral. Universidad Complutense de Madrid, Madrid.

Cortés, O.; Tupac-Yupanqui, I.; Dunner, S.; García-Atance, M.A.; García, D.; Fernández, J.; Cañón, J. (2008) Ancestral matrilineages and mitochondrial DNA diversity of the Lidia cattle breed. *Animal Genetics* (in press).

Crandall, K.A.; Bininda-Emonds, O.R.P.; Mace, G.M.; Wayne, R.K. (2000) Considering evolutionary processes in conservation biology. *Trends in Ecology and Evolution*, 15: 290-295.

Crow, J.F.; Kimura, M. (1970) *An introduction to population genetics theory*. Harper & Row, Publishers, Inc., N.Y.

Cymbron, T.; Loftus, R.T.; Malhiero, M.I.; Bradely, D.G. (1999) Mitochondrial sequence variation suggests an African influence in Portuguese cattle. *Proceedings of the royal Society of London Series B*, 266: 597-603.

Dieringer, D.; Schlotterer, C. (2003) Two distinct modes of microsatellite mutation processes: evidence from the complete genomic sequences of nine species. *Genome Research*, 13: 2242-2251.

Di Rienzo, A.; Peterson, A.C.; Garza, J.C.; Valdes, A.M.; Slatkin, M. et al. (1994) Mutational processes of simple sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences USA*, 91: 3166-3170.

Durand-Tardif, M.; Mitteau, M.; Planchenault (2000) Des clés pour la gestion des ressources génétiques. Quelques acquis du 3e colloque national, Toulouse, 9-11 octobre 2000

EAAP (1998) Assessment of the degree of endangerment of livestock breeds. Working group on Animal Genetic Resources, 49th Annual Meeting European Association of Animal Production, Warsaw, 1998.

Eding, H.; Crooijmans, R.P.M.A.; Groenen, M.A.M.; Meuwissen, T.H.E. (2002) Assessing the contribution of breeds to genetic diversity in conservation schemes. *Genetics, Selection and Evolution*, 34: 613-633.

Eding, H.; Laval, G. (1999) Measuring genetic uniqueness in livestock. In: Oldenbroek JK (ed) *Genebanks and the Conservation of Farm Animal Genetic Resources*. DLO Institute for Animal Sciences and Health: Lelystad, The Netherlands, pp 33-58.

Eding, H.; Meuwissen, T.H.E. (2001) Marker-based estimates of between and within population kinships for the conservation of genetic diversity. *Journal of animal breeding and genetics*, 118: 141-159.

Eding, H.; Meuwissen, T.H.E. (2002) The use of marker estimated kinships to measure genetic diversity. 7th World Congress on Genetics Applied to Livestock Production. August 19-23, Montpellier, France. Comm. 26-03

Eding, H.; Meuwissen, T.H.E. (2003) Linear methods to estimate kinships from genetic marker data for the construction of core sets in genetic conservation schemes. *Journal of Animal Breeding and Genetics*, 120: 289-302.

El Mousadik, A.; Petit, R.J. (1996) High level of genetic differentiation for allelic richness among populations of the argan tree (*Argania spinosa* L. Skeels) endemic of Morocco. *Theoretical and Applied Genetics*, 92: 832-839.

Ellegren, H. (2000) Heterogeneous mutation processes in human microsatellite DNA sequences. *Nature Genetics*, 24: 400-402.

Ellegren, H. (2004) Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics*, 5: 435-445.

Emik, L.O.; Terrill, C.E. (1949) Systematic procedures for calculating inbreeding coefficients. *Journal of Heredity*, 40: 51-55.

Epstein, H. (1971) *The origin of the Domestic Animals of Africa*. Vol.1 (1st ed.) Africana, New York.

Epstein, H., Mason, I.L. (1984) Cattle. In: Mason, I.L., ed., *Evolution of Domesticated Animals*. (1st. Ed.), pp. 6-27. Longman, London.

European Cattle Genetic Diversity Consortium (2006) Marker-assisted conservation of European cattle breeds: an evaluation. *Animal Genetics*, 37: 475-481.

Excoffier, L. (2001) Analysis of population subdivision. In: Balding, D.J.; Bishop, M.; Cannings, C. (ed.) *Handbook of Statistical Genetics*. John Wiley: New York, pp. 271-307.

Fabuel, E.; Barragán, C.; Silió, L.; Rodríguez, M.C.; Toro, M.A. (2004) Analysis of genetic diversity and conservation priorities in Iberian pigs based on microsatellite markers. *Heredity*, 93: 104-113.

Falconer, D.S.; Mackay, T.F.C. (1996) *Introduction to quantitative genetics*. Longman House, Harlow.

FAO (1999) *The Global Strategy for the Management of Farm Animal Genetic Resources*. Executive Brief. FAO, Rome.

Feller, W. (1951) Diffusion processes in genetics. *Second Berkeley Simp. Math. Stat. Probab., Proc.*, 227-246.

Fernández, J.; Toro, M.A.; Caballero, A. (2004) Managing individuals' contributions to maximize the allelic diversity maintained in small, conserved populations. *Conservation Biology*, 18: 1358-1367.

Foulley, J.-L.; Ollivier, L. (2006) Estimating allelic richness and its diversity. *Livestock Science*, 101: 150-158.

Frankham, R. (1994) Conservation of genetic diversity for animal improvement. *Proceedings of the 5th World Congress of Genetics Applied to Livestock Production*, Vol. 21. University of Guelph, Canada, pp. 385-392.

Fraser, D.J.; Bernatchez, L. (2001) Adaptive evolutionary conservation: towards a unified concept for defining conservation units. *Molecular ecology*, 10: 2741-2752.

Gandini, G.; Oldenbroek, K. (2007) Strategies for moving from conservation to utilisation. In Oldenbroek, K., ed.: *Utilisation and conservation of farm animal genetic resources*, pp. 29-54. Wageningen Academic Publishers, The Netherlands.

García, D.; Cañón, J. (2005) Extensiones de la medida de diversidad de Weitzman: inclusión de información intrapoblacional y aproximación del algoritmo. X Conferencia Española de Biometría, Libro de Resúmenes. Ediciones de la Universidad de Oviedo, Oviedo.

García, D.; Corral, N.; Cañón, J. (2005) Combining inter- and intrapopulation information with the Weitzman approach to diversity conservation. *Journal of Heredity*, 96: 704-712.

Gilpin, M.E.; Soulé, M.E. (1986) Minimum viable populations: processes of species extinction. In Soulé, M.E., ed. *Conservation biology: the science of scarcity and diversity*. Sinauer, Sunderland, Massachusetts, USA.

Glover, F.; Kuo, C.-C.; Dhir, K.S. (1995) A discrete optimization model for preserving biological diversity. *Applied Mathematical Modelling*, 19: 696-701.

Goldstein, D.B.; Ruiz Linares, A.; Cavalli-Sforza, L.L.; Feldman, M.W. (1995a) An evaluation of genetic distances for use with microsatellite loci. *Genetics*, 139: 463-471.

Goldstein, D.B.; Ruiz Linares, A.; Cavalli-Sforza, L.L.; Feldman, M.W. (1995b) Genetic absolute dating based on microsatellites and the origin of modern humans. *Proceedings of the National Academy of Sciences USA*, 92: 6723-6727.

Goldstein, D.B.; Pollock, D.D. (1997) Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. *Journal of Heredity*, 88: 335-342.

Goudet, J. (1995) Fstat version 1.2: a computer program to calculate Fstatistics. *Journal of Heredity*, 86: 485-486.

Grigson, C. (1978) The craniology and relationships of four species of *Bos*. IV. The relationship between *Bos primigenius* Boj. and *Bos taurus* L. and its implications for the phylogeny of the domestic breeds. *Journal of Archaeological Science*, 5: 123-152.

Grigson, C. (1980) The craniology and relationships of four species of *Bos*. 5. *Bos indicus* L. *Journal of Archaeological Science*, 7: 3-32.

Gutiérrez, J.P.; Royo, L.J.; Álvarez, I.; Goyache, F. (2005) MolKin v2.0: A computer program for genetic analysis of populations using molecular coancestry information. *Journal of Heredity*, 96: 718-721.

Haig, S.M.; Ballou, J.D.; Derrickson, S.R. (1990) Management options for preserving genetic diversity: reintroduction of Guam rails to the wild. *Conservation Biology*, 4: 290-300.

Hartl, D.L.; Clark, A.G. (1989) *Principles of population genetics* (2nd ed.) Sinauer Associates. Sunderland, Massachusetts.

Hayes, B.; He, J.; Moen, T.; Bennewitz, J. (2006) Use of molecular markers to maximise diversity of founder populations for aquaculture breeding programs. *Aquaculture*, 255: 573-578.

Henderson, S.T.; Petes, T.D. (1992) Instability of simple sequence DNA in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology*, 12: 2749-2757.

Huang, Q.Y.; Xu, F.H.; Shen, H.; Deng, H.Y.; Liu, Y.J.; Liu, Y.Z.; Li, J.L.; Recker, R.R.; Deng, H.W. (2002) Mutation patterns at dinucleotide microsatellite loci in humans. *American Journal of Human Genetics*, 70: 625-634.

Hurlbert, S.H. (1971) The non concept of species diversity: a critique and alternative parameters. *Ecology*, 52: 577-586.

INRA; CIRAD (2002) Contributions to sustainable livelihoods and development: realising sustainable breeding programmes in livestock production. *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production*, Montpellier, 19-23 August. CD-ROM. INRA, Toulouse

Johnston, L.A.; Lacy, R.C. (1995) Genome resource banking for species conservation: selection of sperm donors. *Cryobiology*, 32: 68-77.

Kalinowsky, S.T. (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*, 5: 187-189.

Karlin, S.; Taylor, H.M. (1981) *A first Course in stochastic Processes*. Academic Press, London.

Kelly, R.; Gibbs, M.; Collick, A.; Jeffreys, A.J. (1991) Spontaneous mutation at the hypervariable mouse minisatellites locus Ms6-hm: flanking DNA sequence and analysis of germline and early somatic events. *Proceedings of the Royal Society of London, Series B*, 245: 235-245.

Kimura, M. (1955) Stochastic processes and distribution of gene frequencies under natural selection. *Cold Spring Harbor Symposia on Quantitative Biology*, 20: 33-53.

Kimura, M.; Crow, J.F. (1964) The number of alleles that can be maintained in a finite population. *Genetics*, 49: 725-738.

Kolmogorov, A. (1931) Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. *Mathematische Annalen*, 104: 415-458.

Korn, G.A.; Korn, T.M. (1968) *Mathematical handbook for Scientist and Engineers*, 2nd ed., McGraw-Hill, N.Y.

Latter, B.D.H. (1972a) Selection in finite populations with multiple alleles. III. Genetic divergence with centripetal selection and mutation. *Genetics*, 70: 475-490.

Latter, B.D.H. (1972b) The island model of population differentiation: a general solution. *Genetics*, 73: 147-157.

Latter, B.D.H. (1973) The estimation of genetic divergence between populations based on gene frequency data. *American Journal of Human Genetics*, 25: 247-261.

Laval, G.; Iannuccelli, N.; Legault, C.; Milan, D.; Groenen, M.A.M.; Giuffra, E.; Andersson, L.; Nissen, P.; Jorgensen, C.B.; Beeckmann, P.; Geldermann, H.; Foulley, J-L; Chevalet, C.; Ollivier, L. (2000) Genetic diversity of eleven European pig breeds. *Genetics, Selection and Evolution*, 32: 187-203.

Laval, G.; SanCristobal, M.; Chevalet, C. (2002) Measuring genetic distances between breeds: use of some distances in various short term evolution models. *Genetics, Selection and Evolution*, 34: 481-507

Levinson, G.; Gutman, G.A. (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution*, 4: 203-221.

Loftus, R.T., MacHugh, D.E.; Bradley, D.G.; Sharp, P.M.; Cunningham, P. (1994) Evidence for two independent domestications of cattle. *Proceedings of the National Academy of Sciences USA*, 91: 2757-2761.

Luikart, G.; England, P.R.; Tallmon, D.; Jordan, S.; Taberlet, P. (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Review Genetics*, 4: 981-994.

Lynch, M. (1996) A quantitative-genetic perspective on conservation issues. Pags 471–501 en Avise, J. y Hamrick, J., ed. Conservation genetics: case histories from nature. Chapman and Hall, New York.

Lynch, M. (1998) Estimation of relatedness by DNA fingerprinting. *Molecular Biology and Evolution*, 5: 584-599.

Lynch, M.; Walsh, B. (Draft, 2000) Evolution and selection of quantitative traits.

Malécot, G. (1948) Les mathématiques de l'hérédité. Masson et Cie., Paris.

Manwell, C.; Baker, A.C.M. (1980) Chemical classification of cattle. 2. Phylogenetic tree and specific status of the zebu. *Animal Blood Groups and Biochemical Genetics*, 11: 151-162.

Marshall, D.R.; Brown, H.D. (1975) Optimum sampling strategies in genetic conservation. Pags. 53-80 en Frankel, O.H.; Hawkes, J.G., ed. Crop genetic resources for today and tomorrow. Cambridge University Press, Cambridge, U.K.

Mateus, J.C.; Eding, H.; Penedo, M.C.T.; Rangel-Figueiredo, M.T. (2004) Contributions of Portuguese cattle breeds to genetic diversity using marker-estimated kinships. *Animal Genetics*, 35: 305-313.

MacHugh, D.E.; Shriver, M.D.; Loftus, R.T.; Cunningham, P.; Bradley, D.G. (1997) Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*). *Genetics*, 146: 1071-1086.

McKay, J.K.; Latta, R.G. (2002) Adaptive population divergence: markers, QTL and traits. *Trends in Ecology & Evolution*, 17: 285–291.

Metrick, A.; Weitzman, M.L. (1998) Conflicts and choices in biodiversity preservation. *Journal of Economic Perspectives*, 12: 21-34.

Miretti, M.M.; Dunner, S.; Naves, M.; Contel, E.P.; Ferro, J.A. (2004) Predominant African-derived mtDNA in Caribbean and Brazilian Creole cattle is also found in Spanish cattle. *Journal of Heredity*, 95: 450-453.

Moazami-Goudarzi, K.; Laloë, D.; Furet, J.P.; Grosclaude, F. (1997) Analysis of genetic relationships between 10 cattle breeds with 17 microsatellites. *Animal Genetics*, 28: 338-345.

Moritz, C. (1994) Defining “evolutionarily significant units” for conservation. *Trends in Ecology & Evolution*, 9: 373–375.

Nauta, M.J.; Weissing, F.J. (1996) Constraints on allele size at microsatellite loci: implications for genetic differentiation. *Genetics*, 143: 1021-1032.

Nei, M. (1972) Genetic distance between populations. *American Naturalist*, 106: 283-291.

Nei, M. (1973) The theory and estimation of genetic distance. In: Genetic structure of populations, ed. by N. Morton. University of Hawaii, Honolulu, pp. 45-54.

Nei, M. (1987) *Molecular evolutionary genetics*. Columbia University Press, New York.

Nei M.; Chesser, R.K. (1983) Estimation of fixation indices and gene diversities. *Annals of Human Genetics*, 47: 253-259.

Nei, M.; Tajima, F.; Taten, Y. (1983) Accuracy of estimated phylogenetic trees from molecular data. *Journal of Molecular Evolution*, 19: 153-170.

Nielsen, R.; Palsboll, P.J. (1999) Single-locus tests of microsatellite evolution: multi-step mutations and constraints on allele size. *Molecular Phylogenetics and Evolution*, 11, 477-484.

Ohta, T.; Kimura, M. (1973) A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetical Research Combined*, 22: 201-204.

Oldenbroek, K. (2007) Introduction. In Oldenbroek, K., ed.: *Utilisation and conservation of farm animal genetic resources*, pp. 13-27. Wageningen Academic Publishers, The Netherlands.

Ollivier, L.; Alderson, L.; Gandini, G.C.; Foulley, J-L.; Haley, C.S.; Joosten, R.; Rattink, A.P.; Harlizius, B.; Groenen, M.A.M.; Amigues, Y.; Boscher, M-Y.; Russell, G.; Law, A.; Davoli, R.; Russo, V.; Matassino, D.; Désautés, C.; Fimland, E.; Bagga, M.; Delgado, J-V.; Vega-Pla, J.L.; Martinez, A.M.; Ramos, A.M.; Glodek, P.; Meyer, J-N.; Plastow, G.S.; Siggins, K.W.; Archibald, A.L.; Milan, D.; San Cristobal, M.; Laval, G.; Hammond, K.; Cardellino, R.; Chevalet, C. (2005) An assessment of European pig diversity using molecular markers: Partitioning of diversity among breeds. *Conservation Genetics*, 6: 729-741.

Ollivier, L.; Foulley, J.L. (2002) Some suggestions on how to preserve both within- and between-breed genetic diversity. 53rd Annual Meeting of the European Association for Animal Production. Cairo, Egypt, 1-4 September 2002.

Ollivier, L.; Foulley, J.L. (2004) Objectives in livestock diversity preservation: the European pig example. In *Wissenschaftliches Kolloquium "Nutztierzüchtung im Wandel der Zeit"*, pp. 87-106. Cuvillier Verlag, Göttingen.

Ollivier, L.; Foulley, J.L. (2005) Aggregate diversity: new approach combining within- and between-breed genetic diversity. *Livestock Production Science*, 95: 247-254.

Payne, W.J.A. (1970) *Cattle Production in the Tropics*. (1st ed.) Longman, London.

Payne, W.J.A. (1991) Domestication: a forward step in civilisation. In: Hickman, C.G., ed., *Cattle Genetic Resources*. (1st ed.), pp. 51-72. Elsevier, Amsterdam.

Pääbo, S. (2000) Of bears, conservation genetics, and the value of time travel. *Proceedings of the National Academy of Sciences USA*, 97: 1320-1321.

Petit, R.; El Mousadik, A.; Pons, O. (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology*, 12: 844-855.

- Pike, R.W. (1986) Optimization in Engineering and Systems Science. Van Nostrand Reinhold, N.Y.
- Piyasatian, N.; Kinghorn, B.P. (2003) Balancing genetic diversity, genetic merit and population viability in conservation programmes. *Journal of Animal Breeding and Genetics*, 120: 137-149.
- Pons O.; Chaouche, K. (1995) Estimation, variance and optimal sampling of gene diversity II. Diploid locus. *Theoretical and Applied Genetics*, 91: 122-130.
- Prevosti, A.; Ocana, J.; Alonzo, G. (1975) Distances between populations for *Drosophila subobscura* based on chromosome arrangement frequencies. *Theoretical and Applied Genetics*, 45: 231-241.
- Primmer, C.R.; Ellegren, H.; Saino, N.; Moller, A.P. (1996) Directional evolution in germline microsatellite mutations. *Nature Genetics*, 13: 391-393.
- Reed, C.A. (ed.) (1977) Origins of agriculture. Monton Publishers, The Hague and Paris.
- Reist-Marti, S. B.; Simianer, H.; Gibson, J.; Hanotte, O.; Rege, J.E.O. (2003) Weitzman's approach and conservation of breed diversity: an application to African cattle breeds. *Conservation Biology*, 17: 1299-1311.
- Reist-Marti, S.B.; Abdulaib, A.; Simianer, H. (2006) Optimum allocation of conservation funds and choice of conservation programs for a set of African cattle breeds. *Genetics, Selection and Evolution*, 38: 99-126.
- Renwick, A.; Davison; L.; Spratt, H.; King, J.P.; Kimmel, M. (2001) DNA dinucleotide evolution in humans: fitting theory to facts. *Genetics*, 159: 737-747.
- Reynolds, J.; Weir, B.S.; Cockerham, C.C. (1983) Estimation of the coancestry coefficient: basis for a short term genetic distance. *Genetics*, 105: 767-779.
- Rogers, J.S. (1972) Measures of genetic similarity and genetic distance. In: *Studies in Genetics VII*, University of Texas Publication 7213. University of Texas, Austin, pp. 145-153.
- Roosen, J.; Fadlaoui, A.; Bertaglia, M. (2005) Economic evaluation for conservation of farm animal genetic resources. *Journal of Animal Breeding and Genetics*: 122, 217-228.
- Rosenberg, N.A.; Burke, T.; Elo, K.; Feldman, M.W.; Freidlin, P.J.; Groenen, M.A.M.; Hillel, J.; Mäki-Tanila, A.; Tixier-Boichard, M.; Vignal, A.; Wimmers, K.; Weigend, S. (2001) Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. *Genetics*, 159: 699-713.
- Ruane, J. (1999) A critical review of the values of genetic distance studies in conservation of animal genetic resources. *Journal of Animal Breeding and Genetics*, 116: 317-323.
- Sanghvi, L.D. (1953) Comparison of genetical and morphological methods for a study of biological differences. *American Journal of Physical Anthropology*, 11: 385-404.

- Scherf, B.D. (ed.) (2000) World watch list for domestic animal diversity. 3rd edition. FAO, Rome.
- Shriver, M.D.; Jin, R.; Boerwinkle, E.; Deka, R.; Ferrell, R.E.; Chakraborty, R. (1995) A novel measure of genetic distance for highly polymorphic tandem repeat loci. *Molecular Biology and Evolution*, 12: 914-920.
- Silva, B.; Gonzalo A.; Cañón, J. (2002) Genetic parameters of behavioural traits in the bovine (*Bos taurus*). *Proceedings of the 7th World Congress on Genetics Applied to Livestock*. Vol. XXXII, pp. 83-6.
- Silva, B., Gonzalo, A.; Cañón J. (2006) Genetic parameters of aggressiveness, ferocity and mobility in the fighting bull breed. *Animal Research*, 55: 65-70.
- Simianer, H. (2005) Using expected allele number as objective function to design between and within breed conservation of farm animal biodiversity. *Journal of Animal Breeding and Genetics*, 122: 177-187.
- Simianer, H.; Marti, S.B.; Gibson, J.; Hanotte, O.; Rege, J.E.O. (2003) An approach to the optimal allocation of conservation funds to minimize loss of genetic diversity between livestock breeds. *Ecological Economics*, 45: 377-392.
- Slatkin M (1991) Inbreeding coefficients and coalescence times. *Genetical Research*, 58: 167-175.
- Slatkin, M. (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, 139: 457-462.
- Solis, A.; Jugo, B.M.; Mériaux, J.C.; Iriando, M.; Mazón, L.I.; Aguirre, A.I.; Vicario, A.; Estomba, A. (2005) Genetic diversity within and among four South European native horse breeds based on microsatellite DNA analysis: implications for conservation. *Journal of Heredity*, 96: 670-678.
- Stefanini, F.M.; Feldman, M.W. (2000) Bayesian estimation of range for microsatellite loci. *Genetical Research*, 75: 167-177.
- Stephens, J.C.; Bilbert, D.A.; Yuhki, N.; O'Brien, S.J. (1992) Estimation of heterozygosity for single-probe multilocus DNA fingerprints. *Molecular Biology and Evolution*, 9: 729-743.
- Takezaki, N.; Nei, M. (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics*, 144: 389-399.
- Tapio, I.; Värvi, S.; Bennewitz, J.; Maleviciute, J.; Fimland, E.; Grislis, Z.; Meuwissen, T.H.E.; Miceikiene, I.; Olsaker, I.; Viinalass, H.; Vilkki, J.; Kantanen, J. (2006) Prioritization for conservation of Northern European cattle breeds based on analysis of microsatellite data. *Conservation Biology*, 20: 1768-1779.
- Tautz, D. (1993) Notes on the defunction and nomenclatures of tandemly repetitive DNA sequences, pp. 21-28 in *DNA fingerprinting: State of the Science*, ed. by S.D.J. Pena, R. Chakraborty, J.T. Epplen and A.J. Jeffreys. Birkhäuser Verlag, Basel.
- Thaon d'Arnoldi, C., Foulley, J.L., Ollivier, L., (1998). An overview of the Weitzman approach to diversity. *Genetics, Selection and Evolution*, 30: 149-161.

van Tienderen, P.H.; de Haan, A.A.; van der Linden, C.G.; Vosman, B. (2002) Biodiversity assessment using markers for ecologically important traits. *Trends in Ecology & Evolution*, 17: 577–582.

Toro, M.; Barragán, C.; Óvilo, C.; Rodrigañez, J.; Rodríguez, C.; Silió, L. (2002) Estimation of coancestry in Iberian pigs using molecular markers. *Conservation Genetics*, 3: 309-320.

Toro, M.; Silió, L.; Rodríguez, J.; Rodríguez, C. (1998) The use of molecular markers in conservation programmes of live animals. *Genetics, Selection and Evolution*, 30: 585-600.

Troy, C.S.; MacHugh, D.E.; Bailey, J.F.; Magee, D.A.; Loftus, R.T.; Cunningham, P.; Chamberlain, A.T.; Sykes, B.C.; Bradley, D.G. (2001) Genetic evidence for Near-Eastern origins of European cattle. *Nature*, 410:1088-1091.

Walsh, G. R. (1979) *Methods of Optimization*, John Wiley and Sons, Inc., New York (1979).

Weber, J.; Wong, C. (1993) Mutation in human short tandem repeats. *Human Molecular Genetics*, 2: 1123-1128.

Weikard, H.P. (2002) Diversity Functions and the Value of Biodiversity. *Land Economics*, 78: 20-27.

Weir, B.S. (1996) *Genetic data analysis II*. Sinauer Associates, Sunderland, Massachusetts.

Weir, B.S.; Cockerham, C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, 38: 1358-1370.

Weitzman, M.L. (1992) On diversity. *The Quarterly Journal of Economics*, 107: 363-405.

Weitzman, M. (1993). What to preserve?. An application of diversity theory to crane conservation. *The Quarterly Journal of Economics*, 108: 157-183.

Weitzman, M.L. (1998) The Noah's ark problem. *Econometrica*, 66: 1279-1298.

Wiener, P.; Burton, D.; Ajmone-Marsan, P.; Dunner, S.; Mommens, G.; Nijman, U.; Valentini, A.; Zaragoza, P.; Williams, J-L. (2003) Signatures of selection? Patterns of microsatellite diversity on a chromosome containing a selected locus. *Heredity*, 90: 350-358.

Woolliams, J.; Toro, M. (2007) What is genetic diversity? In Oldenbroek, K., ed.: *Utilisation and conservation of farm animal genetic resources*, pp. 55-74. Wageningen Academic Publishers, The Netherlands.

Wright, S. (1949) Adaptation and selection. In: Jepson, G.L., Simpson, G.G.; Mayr, E., ed. *Genetics, Paleontology, and Evolution* Princeton University Press, Princeton.

Wright, S. (1951) The genetical structure of populations. *Annals of Eugenetics*, 15: 323-354.

Wright, S. (1978) *Evolution and the Genetics of Populations*. Vol. 4. Variability Within and Among Natural Populations. Univ. of Chicago Press, Chicago.

Xu, W.; Humphries, S.; Tomita, M.; Okuyama, T.; Matsuki, M.; Burgner, D.; Kwiatkowski, D.; Liu, L.; Charles, I.G. (2000a) Survey of the allelic frequency of a NOS2A promoter microsatellite in human populations: assessment of the NOS2A gene and predisposition to infectious disease. *Nitric Oxide*, 4: 379-383.

Xu, X.; Peng, M.; Fang, Z. (2000b) The direction of microsatellite mutations is dependent upon allele length. *Nature Genetics*, 24: 396-399.

Zeuner, F.E. (1963) The history of the domestication of cattle. In: Mourant, A.E.; Zeuner, F.E., ed., *Man and Cattle*. Royal Anthropological Institute of Great Britain and Ireland. (vol.18) pp. 9-20.

Zheng, Y.Q.; Lindgren, D.; Rosvall, O.; Westin, J. (1997) Combining genetic gain and diversity by considering average coancestry in clonal selection of Norway spruce. *Theoretical and Applied Genetics*, 95: 1312-1319.

Zhivotovsky, L.A.; Rosenbert, N.A.; Feldman, M.W. (2003) Features of evolution and expansion of modern humans, inferred from genomewide microsatellite markers. *American Journal of Human Genetics*, 72: 1171-1186.

Apéndice: publicaciones relacionadas

Genetic diversity measures of local European beef cattle breeds for conservation purposes

Javier CAÑÓN^{a,*}, Paolo ALEXANDRINO^b,
Isabel BESSA^b, Carlos CARLEOS^c, Yolanda CARRETERO^a,
Susana DUNNER^a, Nuno FERRAN^b, David GARCIA^a,
Jordi JORDANA^d, Denis LALOË^e, Albano PEREIRA^b,
Armand SANCHEZ^d, Katayoun MOAZAMI-GOUDARZI^e

^a Laboratorio de Genética, Facultad de Veterinaria,
Universidad Complutense de Madrid, 28040 Madrid, Spain

^b Faculdade Ciencias Porto, 04021 Porto, Portugal

^c Facultad de Ciencias, Universidad de Oviedo, 33007 Oviedo, Spain

^d Facultad de Veterinaria, Universidad Autónoma de Barcelona,
08193 Bellaterra, Spain

^e Institut national de la recherche agronomique, Département de génétique animale,
78352 Jouy-en-Josas, France

(Received 4 August 2000; accepted 2 January 2001)

Abstract – This study was undertaken to determine the genetic structure, evolutionary relationships, and the genetic diversity among 18 local cattle breeds from Spain, Portugal, and France using 16 microsatellites. Heterozygosities, estimates of F_{st} , genetic distances, multivariate and diversity analyses, and assignment tests were performed. Heterozygosities ranged from 0.54 in the Pirenaica breed to 0.72 in the Barrosã breed. Seven percent of the total genetic variability can be attributed to differences among breeds (mean $F_{st} = 0.07$; $P < 0.01$). Five different genetic distances were computed and compared with no correlation found to be significantly different from 0 between distances based on the effective size of the population and those which use the size of the alleles. The Weitzman recursive approach and a multivariate analysis were used to measure the contribution of the breeds diversity. The Weitzman approach suggests that the most important breeds to be preserved are those grouped into two clusters: the cluster formed by the Mirandesa and Alistana breeds and that of the Sayaguesa and Tudanca breeds. The hypothetical extinction of one of those clusters represents a 17% loss of diversity. A correspondence analysis not only distinguished four breed groups but also confirmed results of previous studies classifying the important breeds contributing to diversity. In addition, the variation between breeds was sufficiently high so as to allow individuals to be assigned to their breed of origin with a probability of 99% for simulated samples.

local beef cattle breeds / microsatellite / genetic diversity

* Correspondence and reprints

E-mail: jcanon@eucmax.sim.ucm.es

1. INTRODUCTION

During the last forty years, it has become clear that biochemical analyses of genetic variation can provide valuable insight into the genetic structure and evolutionary history of cattle populations. Studies have been undertaken on a broad scale to encompass populations not only from different regions of the globe but also at a local level among closely related populations within particular regions [4,18,22,30,33,38]. Manwell and Baker [31] were the first to present a phylogenetic tree for the ten major cattle breed-groups of Europe, Western Asia, and Northern Africa. By reviewing the data on protein polymorphism, they were able to demonstrate that it was in positive agreement with morphological and geographical divisions of the major breed-groups. They were not able, however, to study relationships between individual breeds.

More recently, molecular techniques have provided new markers for the study of genetic variation [6,27,37]. Among these, microsatellites (repetitive elements containing simple sequence motifs, usually dimers or trimers) have quickly become the favourite agents for population genetic studies as they offer advantages which are particularly appropriate in conservation projects. First, they are widely available. Second, they exhibit a high degree of polymorphism. Third, as genetic systems, they are comparatively easy to automate with the possibility of multiplex amplification of up to five loci in a single PCR reaction and of multiple loadings of up to fifteen loci per lane in some highly optimised gel systems. In addition, it is assumed they are neutral to selection, the observed genetic diversity being the consequence of two forces: genetic drift and mutation.

In the last five years, different studies of genetic relationships between cattle breeds using microsatellites have been published. MacHugh *et al.* [28] analysed 20 microsatellites in different cattle populations from Africa, Europe, and Asia highlighting a marked distinction between humpless (taurine) and humped (zebu) cattle which provides strong support for the hypothesis of a separate origin of domesticated zebu cattle. Studies aimed at characterising relationships within the African group [45] or within the European group of cattle breeds have focused on breeds from Italy [10], Spain [32], Belgium [36], the British Isles [29], France [35], and Switzerland [42]. It is difficult, however, to group the data from these studies together in order to clarify the genetic relationships among the major types of cattle because they do not use a common set of microsatellites. For this reason, the FAO has proposed a list of thirty microsatellites for the analysis of genetic diversity in European cattle breeds.

The primary goal of this study is to assess the genetic variation within, and between, breeds and groups of breeds. A secondary aim is to define a diversity measure which will permit the ranking of breeds for conservation purposes thus providing useful information concerning the relative contribution

to genetic diversity of 18 local cattle breeds from Spain, Portugal, and France using 16 microsatellites (15 of which are from the FAO list).

2. MATERIALS AND METHODS

2.1. Cattle breeds

The breeds included in this study (Tab. I) are characterised by a widespread regional distribution, small population size, and ties to traditional production systems.

Regarding their morphological attributes, most of the breeds show pigmentation similar to their wild ancestor, from reddish-brown to brownish-black, with black pigmentation restricted to the extremities (Alistana, Mirandesa, Maronesa, Barrosã, Asturiana de los Valles, Asturiana de la Montaña, Aubrac). In some breeds (Tudanca, Gasconne and Bruna) red pigmentation tends to lighten considerably as the animals age. The most commonly observed variants are solid black (Morucha and Avileña) and red pigmentation (Retinta, Alentejana, Pirenaica, Salers) although a colour-sided (Mertolenga) breed was also found in this study. Most of the breeds included in the project have never been exposed to reproductive technology or other breeding tools related to artificial discriminative mating thus limiting the male and female gene flow between breeds with individual dispersion only at local levels. Nevertheless, the lack of organised studbooks, most of them created recently, for many of the breeds has facilitated a certain degree of genetic introgression between them.

2.2. Sampling of populations

The sampling process is of great importance as it allows us to determine the kind of inferences which can be made. In order to reflect the current genetic composition, individuals can be considered to have been sampled at random within-generation.

Fresh blood collected in a conservative buffer was taken from 50 individuals (25 males and 25 females).

2.3. Genetic loci studied

The 16 microsatellite loci studied were: CSSM 66, ETH 10, ETH 152, ETH 225, ETH 3, HEL 1, HEL 5, HEL 9, ILSTS 005, INRA 023, INRA 032, INRA 035, INRA 037, INRA 005, INRA 063, and TGLA 44. References and primer sequences are described in Table II. TGLA 44 is the only locus not included in the European Concerted Action AIRE2066 list (FAO list).

Table I. Summary statistics for beef cattle breeds used in microsatellite marker analysis of population structure showing geographical location, sample size (N), observed (H_o) and expected (H_e) heterozygosity and average number of alleles per locus (MNA). Standard errors in parentheses.

Breed	Origin of the samples	N	H_o	H_e	MNA
Alistana	Spain	50	0.629 (0.032)	0.681 (0.027)	6.9 (0.8)
Asturiana Montaña	Spain	50	0.652 (0.037)	0.705 (0.034)	6.6 (0.7)
Asturiana Valles	Spain	50	0.656 (0.045)	0.683 (0.042)	7.0 (0.7)
Sayaguesa	Spain	50	0.654 (0.031)	0.707 (0.028)	6.4 (0.6)
Tudanca	Spain	50	0.596 (0.040)	0.651 (0.036)	6.8 (0.8)
Avileña Negra-Ibérica	Spain	50	0.589 (0.043)	0.692 (0.034)	6.9 (0.7)
Bruna del Pirineus	Spain	50	0.619 (0.033)	0.672 (0.030)	7.1 (0.7)
Morucha	Spain	50	0.640 (0.036)	0.709 (0.039)	6.9 (0.7)
Pirenaica	Spain	50	0.543 (0.052)	0.628 (0.037)	5.8 (0.4)
Retinta	Spain	50	0.614 (0.040)	0.693 (0.033)	6.8 (0.6)
Alentejana	Portugal	50	0.622 (0.054)	0.655 (0.052)	5.8 (0.5)
Barrosã	Portugal	50	0.716 (0.037)	0.708 (0.039)	6.7 (0.6)
Maronesa	Portugal	49	0.635 (0.045)	0.664 (0.041)	6.1 (0.6)
Mertolenga	Portugal	50	0.626 (0.039)	0.671 (0.035)	5.9 (0.5)
Mirandesa	Portugal	50	0.625 (0.037)	0.635 (0.026)	5.5 (0.4)
Aubrac	France	50	0.569 (0.043)	0.611 (0.036)	6.2 (0.6)
Gasconne	France	50	0.630 (0.039)	0.708 (0.023)	7.2 (0.6)
Salers	France	50	0.580 (0.046)	0.631 (0.036)	6.1 (0.6)

2.4. DNA extraction and PCR amplification

DNA was extracted using established procedures [20,41] that guarantee long-term stability of DNA samples. Primers and Polymerase Chain Reaction (PCR) conditions are described in Table II. The PCR analysis of microsatellites was carried out by loading onto standard 7% polyacrilamide denaturing gel using silver staining [2] or fluorescent-labelled PCR primer methods through an automated DNA fragment analyser (Applied Biosystem 373 or 377). In order to ensure the compatibility of results from different equipment and laboratories, 3 types of reference DNA were used: Type 1 = reference DNAs ($n = 9$) from the AIRE 2006, Type 2 = reference DNA ($n = 4$) from this project, Type 3 = reference DNA ($n = 2$) from individual laboratories. Moreover, the accurate sizing of allele fragments of these 15 reference DNAs was checked by each of the four laboratories involved in the study. In addition, to ensure the compatibility of results within each laboratory, Type 3 DNAs were used as standards for each loaded gel.

Table II. References, primer sequences, chromosomal location, mean (Hs) and total (Ht) heterozygosity and experimental parameters for 16 microsatellite markers. *(continued on the next page)*

Locus	Reference	Primer sequences (5'–3')	Chrom.	Tm (deg)	MgCl ₂ (mM)	Cycles	Detected Size range (bp)	Hs	Ht
CSSM 66	Steffen <i>et al.</i> [46]	P1: ACA CAA ATC CTT TCT GCC AGC TGA P2: AAT TTA ATG CAC TGA GGA GCT TGG	14	55	1.5	30	209–177	0.826	0.859
ETH 10	Steffen <i>et al.</i> [46]	P1: GTT CAG GAC TGG CCC TGC TAA CA P2: CCT CCA GCC CAC TTT CTC TTC TC	5	55	1.5	30	225–207	0.729	0.760
ETH 152	Fries <i>et al.</i> [14]	P1: TAC TCG TAG GGC AGG CTG CCT G P2: GAG ACC TCA GGG TTG GTG ATC AG	5	55	1.5	30	211–181	0.685	0.737
ETH 225	Fries <i>et al.</i> [14]	P1: GAT CAC CTT GCC ACT ATT TCC T P2: ACA TGA CAG CCA GC6T GCT ACT	9	55	1.5	30	155–131	0.736	0.774
ETH 3	Fries <i>et al.</i> [14]	P1: GAA CCT GCC TCT CCT GCA TTG G P2: ACT CTG CCT GTG GCC AAG TAG G	19	55	1.5	30	133–109	0.723	0.772
HEL 1	Kaukinen and Varvio [23]	P1: CAA CAG CTA TTT AAC AAG GA P2: AGG CTA CAG TCC ATG GGA TT	15	55	1.5	30	117–103	0.641	0.681
HEL 5	Kaukinen and Varvio [23]	P1: GCA GGA TCA CTT GTT AGG GA P2: AGA CGT TAG TGT ACA TTA AC	21	55	1.5	30	171–147	0.736	0.790
HEL 9	Kaukinen and Varvio [23]	P1: CCC ATT CAG TCT TCA GAG GT P2: CAC ATC CAT GTT CTC ACC AC	8	55	1.5	30	169–143	0.750	0.818

Table II. Continued.

Locus	Reference	Primer sequences (5'-3')	Chrom.	T _m (deg)	MgCl ₂ (mM)	Cycles	Detected Size range (bp)	Hs	Ht
ILSTS 005	Brezinsky <i>et al.</i> [7]	P1: GGA AGC AAT GAA ATC TAT AGC C P2: TGT TCT GTG AGT TTG TAA GC	10	55	1.5	30	186–184	0.385	0.409
INRA 023	Vaiman <i>et al.</i> [51]	P1: GAG TAG AGC TAC AAG ATA AAC TTC P2: TAA CTA CAG GGT GTT AGA TGA ACT C	3	55	1.5	30	221–197	0.776	0.865
INRA 032	Vaiman <i>et al.</i> [51]	P1: AAA CTG TAT TCT CTA ATA GCA C P2: GCA AGA CAT ATC TCC ATT CCT TT	11	55	1.5	30	190–166	0.703	0.755
INRA 035	Vaiman <i>et al.</i> [51]	P1: ATC CTT TGC AGC CTC CAC ATT G P2: TTG TGC TTT ATG ACA CTA TCC G	16	55	1.5	30	114–102	0.442	0.488
INRA 037	Vaiman <i>et al.</i> [51]	P1: GAT CCT GCT TAT ATT TAA CCA C P2: AAA ATT CCA TGG AGA GAG AAA C	4	55	1.5	30	144–114	0.628	0.687
INRA 005	Vaiman <i>et al.</i> [50]	P1: CAA TCT GCA TGA AGT ATA AAT AT P2: CTT CAG GCA TAC CCT ACA CC	12	55	1.5	30	147–139	0.624	0.655
INRA 063	Vaiman <i>et al.</i> [51]	P1: ATT TGC ACA AGC TAA ATC TAA CC P2: AAA CCA CAG AAA TGC TTG GAA G	18	55	1.5	30	187–171	0.632	0.654
TGLA 44	George <i>et al.</i> [15]	P1: AAC TGT ATA TTG AGA GCC TAC CAT G P2: CAC ACC TTA GCG ACT AAA CCA CCA	2	55	1.5	30	178–144	0.750	0.811

2.5. Statistical analysis

The BIOSYS-1 package [47] was used to compute allele frequencies by direct counting, as well as the number of alleles, and unbiased estimates for expected (H_e) and observed (H_o) heterozygosity.

Different genetic distances clustered into three groups were used: 1) genetic distances considered appropriate under a pure drift model where genetic drift was assumed to be the main factor in genetic differentiation among closely related populations or for short-term evolution [39, 48, 52] – using the traditional differentiation-between-population estimator F_{ST} [55] and the Reynolds genetic distance estimator [39]; 2) genetic distances that assume a step-wise-mutation model, *i.e.*, average squared distance [16] and delta-mu squared distance [17]; 3) a non-metric genetic distance based on the proportion of shared alleles [5]. All genetic distances were estimated using MICROSAT [34] except for the Reynolds distance for which the PHYLIP package [13] was used. The product-moment correlation (r) and Mantel test statistic were computed for pairwise comparisons of distance matrices.

After defining groups of breeds by country or by trunk (a set of breeds with a hypothetical common ancestor) using *a priori* information, a hierarchical analysis of variance was carried out which permitted the partitioning of the total genetic variance into components due to inter-individual differences on the one hand and inter-breed differences on the other. Variance components were then used to compute fixation indices [55] and their significance tested using a non-parametric permutation approach described by Excoffier *et al.* [12]. Computation was carried out using the AMOVA (Analysis of Molecular Variance) programme implemented in the ARLEQUIN package [43].

2.5.1. Multivariate correspondence analysis

Phylogenetic reconstruction and the use of genetic distances do not take into account the effects of admixtures between branches. Alternatively, the representation of genetic relationships among a group of populations may be obtained using multivariate techniques which can condense the information from a large number of alleles and loci into a few synthetic variables.

Correspondence Analysis [3, 26] is a multivariate method analogous to the Principal Components analysis but which is appropriate for categorical variables and leads to a simultaneous representation of breeds and loci as a cloud of points in a metric space. As with the Principal Components analysis, axes, which are ranked according to their fraction of information, span this space with each axis independent of the others. Inertia, or dispersion, measures this information, *i.e.*, the direction of maximum inertia is the direction in which the cloud of points is the most scattered. The basic concept of inertia can be related to the well-established population parameter F_{ST} [19] as well as to genetic diversity [24].

Allele frequencies of all loci were used as variables to spatially cluster the breeds using a correspondence analysis based on Chi-square distances to judge proximity between them.

2.5.2. Computing diversity

Following the Weitzman approach [53,54], the Reynolds genetic distances were used to compute marginal losses of genetic diversity. After transforming the genetic distance matrix into a distance matrix with ultrametric properties, a maximum likelihood tree was drawn using NTSYS [40].

2.5.3. Breed assignment

The assignment of an anonymous animal i to a set of breeds, r_1, \dots, r_n , was based on the maximum likelihood discriminate rule, *i.e.*, animal i was assigned to the population which maximises the conditional probability ($P[i|r]$). Let $\hat{P}_{r,l,a}$ be the frequency of allele a in the l locus and r breed, then $P[i|r] = \prod_l h(i, l) \hat{P}_{r,l,a_{il1}} \hat{P}_{r,l,a_{il2}}$, where $h(i, l) = 1$ if $a_{il1} = a_{il2}$ and $h(i, l) = 2$ if $a_{il1} \neq a_{il2}$. When one allele was missing in a specific population, we assigned a small, but positive, probability of the allele in this breed ($1/(2n+1)$) where n was the sample size of the breed [44]. A traditional way of expressing the significance of a particular result is by using the log of likelihood ratio (LOD). If the interest is to classify an anonymous sample in one of two possible populations, it is necessary to determine the distribution of the appropriate statistic under the null hypothesis (H_0) by bootstrap or by simulating allele frequencies. Given that it is not possible to directly determine the LOD distribution when many loci are used, we simulated 100 000 genotypes per breed using allele frequencies according to the assumptions of Hardy-Weinberg and linkage equilibrium. The frequency at which each animal was correctly assigned to its breed provided the probability of assignment, and the distribution of the LODs for pairs of breeds, or populations, allowed for the construction of confidence thresholds.

3. RESULTS

3.1. Variation within, and among, populations

A total of 173 distinct alleles were detected across the 16 loci analysed. The mean number of alleles (MNA) per locus per breed was 6.5 (Tab. I).

Observed and expected heterozygosities per breed ranged from 0.54 (Pirenaica) to 0.72 (Barrosã), and from 0.61 (Aubrac) to 0.71 (Asturiana de Montaña, Barrosã, Morucha and Sayaguesa) respectively (Tab. I).

Levels of apparent breed differentiation were considerable with multilocus F_{ST} values indicating that around 7% of the total genetic variation corresponded to differences between breeds while the remaining 93% corresponded to differences among individuals.

Table III presents F_{ST} values when breeds were considered in pairs. Genetic differentiation values among breeds ranged from 3% for the Aubrac-Salers pair to 15% for the Mirandesa-Tudanca pair. All values were different from 0 ($P < 0.01$). Values above the diagonal in Table III represent the number of individuals between populations exchanged per generation (Nm , where N is the total effective number of animals and m the migration rate) which balanced the diversifying effect of the genetic drift.

The AMOVA analysis permitted the partitioning of the genetic variability between different sources of variation – hypothetical trunks, or countries – and breeds were the main factors in the analysis carried out in this study. Results of the analysis of variance are shown in Table IV. Clearly, variability (excluding individual variability) was taken into account when looking at the breed factor leaving a low, yet significant, genetic variability ($< 1.5\%$) at the trunk (Tab. IVa), or country level (Tab. IVb). Less than 1.5 per cent of the total genetic differences detected was due to the hypothetical trunk (1.43) or to the country of origin (1.36) to which the breeds were assigned.

3.2. Correspondence Analysis

The first two axes contribute 14% and 13% of the total inertia respectively (Fig. 1). The Sayaguesa breed was isolated from the others and represents 12% of the total inertia respective to the other 18 breeds. Axis 1 separates the Mirandesa and Alistana breeds as well but shows no special proximity between the two. Axis 2 separates two blocks: block I (Gasconne, Salers, Aubrac, Bruna) and Block II (Mirandesa, Alistana, Sayaguesa).

The most important alleles are INRA 032 (170 bp) which contributes 17% in Axis 1 and 9% in Axis 2, and ETH 3 (109 bp) which contributes 8% and 6% in Axis 1 and 2, respectively. Allele INRA 032 (170 bp) is a nearly unique characteristic of the Sayaguesa breed with a frequency of 40% that was absent in the other breeds except the Gasconne and Salers (4% and 1%, respectively). Although this allele appeared in only 9% of the entire breed population studied, allele ETH 3 (109 bp) can be closely associated with the Alistana and Mirandesa breeds which demonstrated a 34% and 58% frequency, respectively.

Observing the importance of allele INRA 032 (170 bp), the analysis was repeated excluding this microsatellite, enabling us to detect a change in the axes – a 15% change in the first axis separating the Alistana and Mirandesa from the other breeds and an 11% change in the second axis separating the Sayaguesa from the others. It became clear at this point that inertia, explained by the

Table III. F_{ST} estimates (below diagonal) as a measure of genetic distance between bovine breeds and the number of effective migrants per generation (Nm) (above diagonal) in balance with genetic drift ($F_{ST} = \frac{1}{4Nm + 1}$) (Wright [55]).

	Alistana	AsturMont	AsturVall	Sayaguesa	Tudanca	Alentejana	Barrosa	Maronesa	Mertolenga	Mirandesa	Aubrac	Gasconne	Salers	Avilena	Bruna	Morucha	Pirenaica	Retinta
Alistana																		
AsturMont	0.0502																	
AsturVall	0.0546	0.0321																
Sayaguesa	0.0863	0.0621	0.0727															
Tudanca	0.0803	0.0587	0.0680	0.0718														
Alentejana	0.0659	0.0465	0.0481	0.1053	0.0909													
Barrosa	0.0649	0.0394	0.0579	0.0668	0.0612	0.0575												
Maronesa	0.0746	0.0421	0.0613	0.0778	0.0787	0.0652	0.0450											
Mertolenga	0.0670	0.0442	0.0510	0.0862	0.0875	0.0385	0.0430	0.0570										
Mirandesa	0.0548	0.0863	0.0959	0.1311	0.1478	0.0976	0.1035	0.1114	0.0901									
Aubrac	0.0957	0.0673	0.0726	0.0990	0.0898	0.0930	0.0984	0.0962	0.0778	0.1207								
Gasconne	0.0783	0.0421	0.0490	0.0822	0.0732	0.0778	0.0673	0.0680	0.0635	0.1062	0.0421							
Salers	0.0868	0.0588	0.0572	0.0967	0.0912	0.0939	0.0986	0.0775	0.0715	0.1204	0.0286	0.0421						
Avilena	0.0581	0.0528	0.0470	0.0758	0.0731	0.0563	0.0635	0.0697	0.0577	0.0829	0.0520	0.0549	0.0576					
Bruna	0.0621	0.0502	0.0422	0.0889	0.0719	0.0708	0.0687	0.0781	0.0615	0.1159	0.0722	0.0515	0.0587	0.0466				
Morucha	0.0694	0.0360	0.0367	0.0759	0.0550	0.0477	0.0449	0.0594	0.0490	0.1077	0.0656	0.0476	0.0608	0.0453	0.0417			
Pirenaica	0.0741	0.0577	0.0468	0.1100	0.0991	0.0747	0.0875	0.0828	0.0564	0.1184	0.0596	0.0386	0.0558	0.0603	0.0398	0.0616		
Retinta	0.0639	0.0434	0.0617	0.0663	0.0489	0.0516	0.0494	0.0386	0.0914	0.0797	0.0472	0.0692	0.0542	0.0565	0.0478	0.0593	0.0670	

Table IV. Partitioning of genetic variability among the different sources of variation.

(a)

Source of variation	Degree of freedom	Sum of squares	Variance components	Percentage of variation	Fixation indices
Among trunks ⁽¹⁾	5	311.13	0.083	1.43	$F_{SC} = 0.057$
Among breeds within trunks	12	455.98	0.327	5.65	$F_{ST} = 0.071$
Within breeds	1 780	9 557.8	5.37	92.9	$F_{CT} = 0.014$
Total	1 797	10 325.8	5.78		

⁽¹⁾ The following 6 arbitrary trunks were defined: (Alistana AsturMont AsturVall Tudanca Sayaguesa); (Mirandesa Barrosã Maronesa); (Aubrac Gasconne Salers); (Bruna Pirenaica); (Retinta Alentejana Mertolenga); (Avilena Morucha).

(b)

Source of variation	Degree of freedom	Sum of squares	Variance components	Percentage of variation	Fixation indices
Among countries	2	163.3	0.079	1.36	$F_{SC} = 0.061$
Among breeds within countries	15	603.8	0.35	6.02	$F_{ST} = 0.074$
Within breeds	1 780	9 558.7	5.37	92.6	$F_{CT} = 0.014$
Total	1 797	10 325.8	5.78		

change from 12% to 7.2% in the Sayaguesa breed, no longer discriminated this breed from the rest since, for example, the Mirandesa had an inertia of 9.4%.

In summary, the Sayaguesa is a breed which can be differentiated from the others, however, this result was obviously amplified by the presence of allele INRA 032 (170 bp) which was present in 40% of the breed and absent, or rare, in the other breeds. Taking into account the position of the Sayaguesa breed, we repeated the analysis excluding this breed. This caused a radical change in the results, which created a zooming-in effect on the other 17 breeds and thus facilitated our ability to interpret the findings.

In this case, Axis 1 explains 16% of the inertia and separates Block 1 (Gasconne, Salers, Aubrac, Pirenaica and Bruna) from Block 2 (Alistana and Mirandesa). The alleles which contributed the most in this axis were INRA 032 (170 bp) (12% contribution) and INRA 037 (126 bp) (6% contribution), the latter having a mean frequency of 17%. INRA 037 (126 bp) could also be found in the Alistana and Mirandesa breeds with frequencies of 41 and 54% respectively though these frequencies were much lower in the Gasconne (4%), Salers (2%), Aubrac (3%), and Pirenaica (11%) breeds. Axis 2 explains 11%

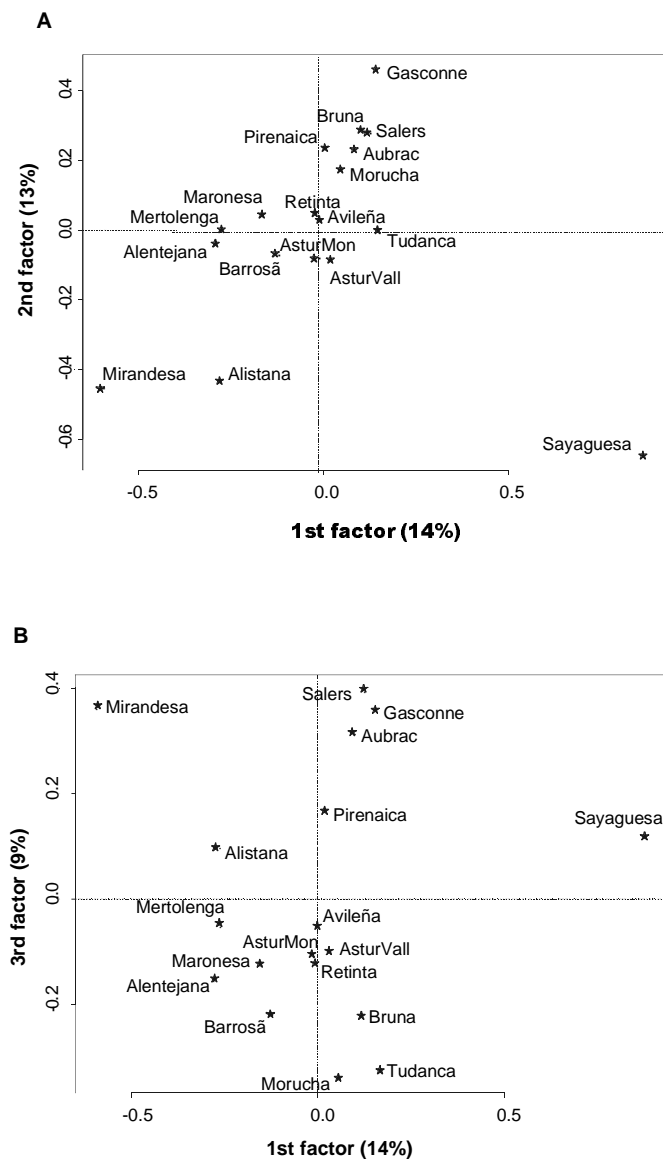


Figure 1. Correspondence analysis of allele frequencies from 16 microsatellite loci typed in eighteen bovine breeds: A) Projection of breeds on axes 1 and 2, B) Projection of breeds on axes 1 and 3.

of the inertia and separates the Morucha, Tudanca, and Bruna block from the Gasconne, Salers, Aubrac and Pirenaica (excluding the Bruna) group. Axis 3 explains 10% of the inertia and singles out the Mertolenga, Barrosã, Maronesa and Alentejana block.

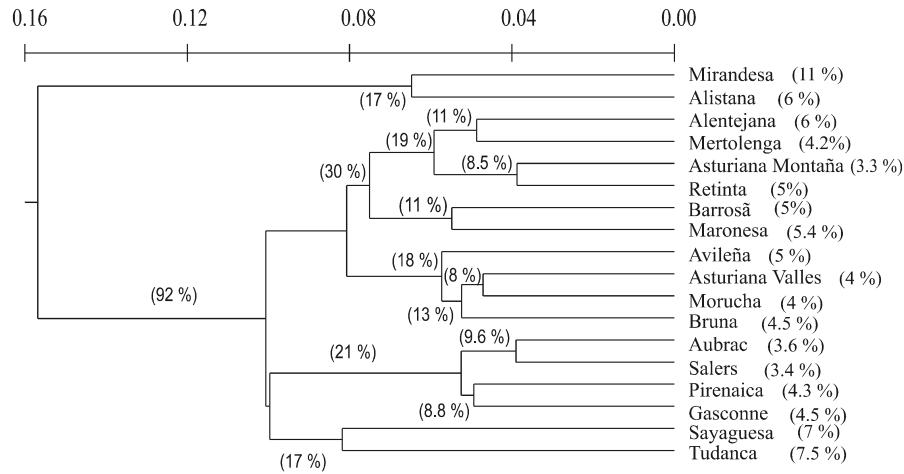


Figure 2. Tree of relationships among 18 local cattle breeds. Values in brackets represent the loss of diversity caused by the extinction of a breed or a set of breeds.

3.3. Evaluation of diversity

In contrast to traditional hierarchical clustering methods, the use of the concepts of link and representative elements (breeds) allows for a unique topology [49]. The tree generated by the algorithm (Fig. 2) has the property of a maximum evolutionary likelihood and the diversity function defined is equal to the total branch length of the tree. The loss of diversity caused by the extinction of a breed, or a set of breeds, can be approximately inferred by looking at the tree or can be exactly quantified by recalculating the total amount of diversity after eliminating the breed, or set of breeds, in question. For instance, a value of 11 585 was found when computing the diversity of the initial set of breeds, and it dropped down to 10 712, a 17% loss of diversity, after the elimination of the Sayaguesa and Tudanca breeds.

3.4. Breed assignment

Results for the assignment of animals to populations using 16 microsatellites are presented in Table V, where the assignment of 100 000 simulated individuals to the breeds is shown. Misclassified individuals were distributed among all breeds. The Sayaguesa and Mirandesa were the breeds most often correctly classified, and the Retinta and Barrosã those most frequently misclassified. Apart from the Salers breed in which 50% of the misclassified individuals were assigned to the Aubrac breed, we did not observe any systematic assignment of animals from one breed to another.

The set of markers used in this study provided a high discriminant power between pairs of breeds: for two closely related populations as are both

Table V. Breed assignment using 16 microsatellites and the maximum likelihood classification rule for eighteen bovine breeds.

	Alistana	Asturleon	AsturVall	Sayaguesa	Tudanca	Alentejana	Barrosa	Maronesa	Mertolenga	Mirandesa	Aubrac	Gasconne	Salers	Avileña	Bruna	Morucha	Pirenaica	Retinta
Alistana	97.7	0.3	0.5	0.03	0.3	0.04	0.2	0.04	0.00	0.6	0.02	0.00	0.05	0.06	0.1	0.04	0.08	0.06
AsturMon	0.3	95.2	0.5	0.04	0.2	0.5	0.6	0.2	0.2	0.1	0.2	0.02	0.2	0.2	0.2	0.4	0.2	0.7
AsturVall	0.7	0.5	94.5	0.1	0.3	0.5	0.5	0.2	0.2	0.1	0.05	0.1	0.2	0.4	0.4	0.4	0.3	0.4
Sayaguesa	0.05	0.03	0.2	99.4	0.05	0.04	0.04	0.00	0.02	0.00	0.00	0.01	0.01	0.05	0.01	0.02	0.02	0.06
Tudanca	0.2	0.3	0.2	0.02	97.6	0.06	0.3	0.07	0.02	0.01	0.1	0.02	0.07	0.1	0.1	0.5	0.1	0.2
Alentejana	0.09	0.2	0.1	0.00	0.04	96.8	0.2	0.1	0.7	0.2	0.06	0.06	0.01	0.2	0.04	0.5	0.1	0.4
Barrosa	0.3	0.5	0.6	0.02	0.3	0.7	93.9	1.1	0.8	0.2	0.02	0.2	0.00	0.3	0.1	0.4	0.2	0.4
Maronesa	0.1	0.2	0.2	0.02	0.1	0.3	1.7	95.6	0.4	0.1	0.03	0.1	0.07	0.2	0.04	0.3	0.2	0.2
Mertolenga	0.01	0.1	0.08	0.00	0.04	0.8	0.6	0.3	96.4	0.2	0.2	0.08	0.04	0.08	0.02	0.06	0.5	0.5
Mirandesa	0.3	0.02	0.06	0.00	0.00	0.1	0.1	0.04	0.3	98.9	0.03	0.01	0.00	0.05	0.00	0.01	0.03	0.06
Aubrac	0.02	0.2	0.04	0.01	0.08	0.1	0.06	0.01	0.2	0.03	95.1	0.7	2.1	0.3	0.1	0.1	0.8	0.08
Gasconne	0.01	0.3	0.2	0.01	0.05	0.1	0.2	0.09	0.03	0.00	0.9	95.2	0.8	0.2	0.2	0.3	1.1	0.4
Salers	0.06	0.1	0.1	0.01	0.08	0.01	0.02	0.01	0.03	0.02	1.8	0.5	96.4	0.08	0.2	0.06	0.4	0.1
Avileña	0.2	0.1	0.3	0.05	0.1	0.6	0.2	0.2	0.2	0.1	0.4	0.3	0.2	95.1	0.3	0.4	0.4	0.5
Bruna	0.1	0.2	0.3	0.01	0.2	0.04	0.01	0.03	0.03	0.02	0.06	0.1	0.3	0.2	96.5	1.0	0.7	0.2
Morucha	0.05	0.4	0.2	0.01	0.2	0.5	0.3	0.5	0.5	0.02	0.2	0.2	0.08	0.2	1.0	95.7	0.2	0.4
Pirenaica	0.08	0.07	0.2	0.00	0.04	0.2	0.03	0.05	0.5	0.08	0.9	0.7	0.5	0.1	0.6	0.2	95.3	0.5
Retinta	0.1	0.6	0.2	0.03	0.2	0.7	0.3	0.4	0.8	0.1	0.1	0.3	0.08	0.6	0.2	0.5	0.6	94.2

Asturiana breeds, only 1.2% of the individuals were misclassified. This can be interpreted from a classical hypothesis testing point of view; if for a certain anonymous sample the test " H_0 : the sample is *Asturiana de Valles*, H_1 : the sample is *Asturiana de Montaña*" is carried out and we set a conservative significance level (0.01), the power of the test ($1 - \text{Pr}(\text{type II error})$) is 0.98.

4. DISCUSSION

Assuming that we are working with a neutral polymorphism, three forces remain that can be used to explain the genetic diversity observed: mutation, genetic drift, and migration. Since mutation is important only when studying long periods of time, we accept that the forces to be considered in this sort of study are genetic drift, the source which contributes to diversity, and migration, the opposite force which tends to homogenise the breeds. Reproductive isolation, a consequence of the local use and management of a breed, reduces the effective population size and contributes to a genetic subdivision that can be detected through drift-based measures based on variations observed when using the microsatellite loci.

The degree of genetic differentiation among the breeds studied and the high levels of significance for the between-population F_{ST} estimates indicate a relatively low gene flow between these breeds and, equivalently, a relatively high reproductive isolation. It is also clear that most of the genetic variation is inter-individual and only less than seven percent of the total variation is due to breed differences.

Migration values (Nm) can be interpreted in the context of the conservation and maintenance of the genetic variability of an animal as the upper limit of the number of migrants per generation which would allow for maintenance of the genetic differentiation observed between the breeds.

Although ancestral trunks are evident in studies based on morphological traits, *e.g.* Jordana *et al.* [21], they are not nearly as apparent when using neutral information to assign breeds to clusters such as the Brown trunk (both Asturian breeds, Alistana, Sayaguesa, Tudanca, etc.), Turdetanus trunk (Pirenaica and Bruna), or Iberian trunk (Avileña and Morucha). Results of this study are confusing since a similar magnitude of differentiation was found among breeds within a trunk or country (5.7% and 6.1% respectively). F_{SC} and F_{ST} are measures of the degree of resemblance between individuals within a breed. This resemblance can be interpreted as the differences between individuals in different breeds and expressed as the differences between breeds as a proportion of the total genetic variance (F_{ST}) or as a proportion of the trunk or country variance (F_{SC}). Conversely, the parameter F_{CT} is a measure of the degree of resemblance between individuals of a trunk, or country, expressed as a proportion of the total variance. The degree of genetic differentiation among

Table VI. Randomized Mantel test statistic (*Z*) for distance matrix comparison.

	F_{ST}	Allele sharing	Deltamu	Average ^(a)
Reynolds	0.999 ^(b)	0.918 ^(b)	0.304	−0.215
F_{ST}		0.917 ^(b)	0.298	−0.223
Allele sharing			0.359	−0.142
Deltamu				0.547

^(a) Average squared distance.

^(b) $\Pr[\text{random } Z \geq \text{observed } Z] < 0.01$.

breeds of different trunks, or countries, was 7.1 and 7.4% respectively, values which are very close to the global degree of genic differentiation among breeds ($F_{ST} = 6.8$) and which clearly show the small genetic contribution the trunk or country factors make.

The lack of correlation between the group of genetic distance measurements which apply under a classical random drift-mutation model and the group which applies under the pure drift model (Tab. VI) is a consequence of the nature of the populations included in this study which cannot be considered as separate, closed populations. European cattle breeds must be considered to be closely related and the main factor describing their genetic variability is random drift. Under this assumption, genetic distances which reflect only the consequences of the genetic drift such as the F_{ST} and Reynolds distances can be considered the most appropriate in measuring the degree of diversification [11], though they could also be inferred comparing the heterozygosity values found with the effective sizes of the breeds, which ranged from 21 (Sayaguesa) (Cañón, personal communication) to over 1 400 (Aubrac, Gasconne) (Renand, personal communication) [25].

Regarding the correspondence analysis, it should be noted that the most significant result was the very strong separation of the Sayaguesa, though this was dependent on the presence of a special allele. This result is obviously not very robust. A very distinct clade is the Gasconne, Pirenaica, Salers and Aubrac block. An Alistana and Mirandesa block was easily distinguished as well even though these two breeds were not very close to one another. Finally, there is the Mertolenga, Barrosa, Alentejana and Maronesa block, though it is less homogeneous than the two cited above.

Looking at Figure 2, where the contribution of each breed to diversity and clade is represented, it is clear that the reduction in diversity as a consequence of the extinction of a clade equals the sum of the reductions caused by the extinction of the breeds which composed the clade. This additive property occurs only if breeds are independent, *e.g.*, the loss of the Mirandesa and Alistana has this property. However, the joint extinction of the Sayaguesa and

Tudanca breeds reduces the total diversity by a greater magnitude than the sum of the two, so they cannot be considered as independent from each other. An interesting question is to what extent both procedures, correspondence analysis and the Weitzman approach, give similar results. It must be emphasised that a correspondence analysis exploits within-breed variability while the Weitzman approach does not. The correlation between the contribution of breeds to diversity, computed by the Weitzman procedure, and the correspondence analysis (inertia) when the complete set of 16 markers and INRA 032 were eliminated, was 0.54 and 0.64 ($p < 0.05$), respectively. Moreover, if we consider the four breeds which contributed the most to diversity, three of them (Mirandesa, Sayaguesa and Alistana) were always present, independently of the analysis procedure used.

Two additional considerations with respect to the Weitzman diversity function refer to the caution needed when interpreting the graphical representation as a phylogenetic tree. Indeed it is only a representation of the diversity found at the current time and the sensitivity of the graphical representation from the model used to study the divergence among the breeds. The order of the breeds appearing in the tree strongly depends on the force (random drift or mutation) considered to be the determinant of the observed diversity. When F_{ST} and Reynolds genetic distances were used, breeds ranked in a similar order (Spearman correlation = 1.0); however, no rank correlation was found to be significantly different from 0 between the breed-order computed using former distances based on effective population size and the breed-order calculated using those genetic distances which are based on the size of the alleles. It should be noted that, despite criticism of the Weitzman approach [49], it continues to be a valid method of determining priorities for conservation investments, if we know the relationships of breeds to each other, the survival probability distribution functions and the costs of improving breed survival.

A different argument showing that hypervariable microsatellites with a high level of heterozygosity and a large number of alleles, provide an efficient way of evaluating genetic diversity between the bovine breeds considered, can be demonstrated by observing their statistical power for breed-affiliation estimation. The results presented in Table V demonstrate the possibility of assigning breed identities to anonymous bovine samples as has been previously shown in equines [9], cattle [29], sheep [8] and humans [44]. These molecular markers provide a powerful tool for measuring the genetic differentiation between breeds of domestic species.

5. CONCLUSIONS

The main objective of conservation genetics is to preserve variability within populations under the hypothesis of correlation between genetic variation and

population viability. Avoidance of inbreeding has often been considered as synonymous with heterozygosity maintenance. Heterozygosity is retained through the maximisation of the inbreeding effective size, which primarily depends on the parental generation size. In populations with known pedigrees, as is the case in this study, maximising effective size while ignoring the ancestry of each individual may not be the most effective strategy for maintaining genetic diversity. Instead, a strategy that utilises all pedigree information would better serve to preserve genetic variation. Unfortunately, many of the local breeds included in this study have incomplete pedigrees and one or both parents of some individuals are unknown. In this context, the application of molecular information can solve some of the uncertainties since it is useful when identifying pedigree relationships and the genetically most important animals in order to maximise founder genome equivalents. Moreover, although additional information on productive, morphological, and fitness-related traits should be taken into account when ranking breeds for preservation purposes, strategies based on neutral markers can be efficient in maximising the retention of the highest number of neutral and non-neutral alleles in small populations [1].

This study contributes to the knowledge of the genetic diversity across different countries and to the molecular characterisation of limited-size populations, many of which are under threat of extinction. It also shows how microsatellites can be used to construct an appropriate measure of diversity function through the genetic relationships between populations. Additionally, the present study provides reasonable statistical power for breed assignment regardless of whether breeds are closely related or not. These issues allow for future management of the breeds to be based on greater knowledge of their genetic structuring and the relationships between their populations.

ACKNOWLEDGEMENTS

This study received the financial support of the EC DGVI FII.3 (contract FAIR1 CT95 0702). Additional support was provided by CICYT (contract AGF96–1950–CE/95). Blood samples from the Sayaguesa and Alistana breeds were provided by E. Matorra and J.E. Yanes, and from the Tudanca breed by R. Sainz. We would also like to thank ASEAVA, ASEAMO, UNION AUBRAC, UPRA SALERS, UPRA GASCONNE and Alain Havy from the Institut de l'Élevage for their help in blood sampling.

REFERENCES

- [1] Bataillon T.M., David J.L., Schoen D.J., Neutral genetic markers and conservation genetics: simulated germplasm collections, *Genetics* 144 (1996) 409–417.

- [2] Bassam B.J., Caetano-Anolles G., Gresshoff P.M., Fast and sensitive silver staining of DNA in polyacrylamide gels, *Anal. Biochem.* 196 (1991) 80–83.
- [3] Benzécri J.P., *Analyse des données. Tome 2: Analyse des correspondances*, Dunod, Paris, 1973.
- [4] Blott S.C., Williams J.L., Haley C.S., Genetic relationships among European cattle breeds, *Anim. Genet.* 29 (1998) 273–282.
- [5] Bowcock A.M., Ruiz-Linares A., Tomforhde J., Minch E., Kidd J.R., Cavalli-Sforza L.L., High resolution of human evolutionary trees with polymorphic microsatellites, *Nature* 368 (1994) 455–457.
- [6] Bradley D.G., Machugh D.E., Loftus R.T., Sow R.S., Hoste C.H., Cunningham E.P., Zebu-taurine variation in Y chromosomal DNA: a sensitive assay for genetic introgression in west African trypanotolerant cattle populations, *Anim. Genet.* 25 (1994) 7–12.
- [7] Brezinsky L., Kemp S.J., Womack J., Teale A.J., A panel of bovine microsatellite genetic markers, in: *Proceedings of the XXII Conference on Animal Genetics*, 1992, Interlaken, Switzerland, International Society for Animal Genetics.
- [8] Buchanan F.C., Adams L.J., Littlejohn R.P., Maddox J.F., Crawford A.M., Determination of evolutionary relationships among sheep breeds using microsatellites, *Genomics* 22 (1994) 397–403.
- [9] Cañon J., Checa M.L., Carleos C., Vega-Pla J.L., Vallejo M., Dunner S., The genetic structure of Spanish Celtic horse breeds inferred from microsatellite data, *Anim. Genet.* 31 (2000) 39–48.
- [10] Ciampolini R., Moazami-Goudarzi K., Vaiman D., Dillman C., Mazzanti E., Foulley J.L., Leveziel H., Cianci D., Individual multilocus genotypes using microsatellite polymorphisms to permit the analysis of the genetic variability within and between Italian beef cattle breeds, *J. Anim. Sci.* 73 (1995) 3259–3268.
- [11] Eding J.H., Laval G., Measuring genetic uniqueness in livestock, in: J.K. Oldenbroek (Ed.), *Genebanks and the conservation of farm animal genetic resources*, DLO Inst. Anim. Sci. and Health, Lelystad, 1999.
- [12] Excoffier L., Smouse P.E., Quattro J.M., Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data, *Genetics* 131 (1992) 479–491.
- [13] Felsenstein J., PHYLIP (Phylogeny Inference Package) Version 3.5c. Department of Genetics, University of Washington, Seattle, 1993.
- [14] Fries R., Eggen A., Womack J.E., The bovine genome map, *Mamm. Genome* 4 (1993) 405–28.
- [15] Georges M., Massey J., Polymorphic DNA markers in bovidae (World Intellectual Property Org Geneva), *Wo. Publi. N° 92/13102*, 1992.
- [16] Goldstein D.B., Ruiz-Linares A., Cavalli-Sforza L.L., Feldman M.W., An evaluation of genetic distances for use with microsatellite loci, *Genetics* 139 (1995) 463–471.
- [17] Goldstein D.B., Ruiz-Linares A., Cavalli-Sforza L.L., Feldman M.W., Genetic absolute dating based on microsatellites and the origin of modern humans, *Proc. Nat. Acad. Sci. USA.* 92 (1995) 6723–6727.
- [18] Grosclaude F., Aupetit R.Y., Lefebvre J., Mériaux J.C., Essai d'analyse des relations génétiques entre les races bovines françaises à l'aide du polymorphisme biochimique, *Genet. Sel. Evol.* 22 (1990) 317–338.

- [19] Guinand B., Use of a multivariate model using allele frequency distributions to analyse patterns of genetic differentiation among populations, *Biol. J. Linn. Soc.* 58 (1996) 173–195.
- [20] Jeanpierre M., A rapid method for purification of DNA from blood, *Nucleic Acids Res.* 15 (1987) 9611.
- [21] Jordana J., Pelegrin M., Piedrafito J., Relaciones genéticas en bovinos españoles obtenidas a partir del estudio de caracteres morfológicos, *ITEA* 87 (1991) 50–64.
- [22] Kantanen J., Olsaker I., Adalsteinsson S., Sandberg K., Eythorsdottir E., Pirhonen K., Holm L.E., Temporal changes in genetic variation of North European cattle breeds, *Anim. Genet.* 30 (1999) 16–27.
- [23] Kaukinen J., Varvio S.L., Eight polymorphic bovine microsatellites, *Anim. Genet.* 24 (1993) 148.
- [24] Laloë D., Moazami-Goudarzi K., Souvenir Zafindrajaona P., Analyse des correspondances et biodiversité dans les races domestiques, *Société française de biométrie*, 20 mai 1999, Grenoble, 5 p.
- [25] Laval G., Iannuccelli A., Legault Ch., Milan D., Groenen A.M., Giuffra E., Andersson L., Nissen P.H., Jorgensen C.B., Biecekman P., Geldermann H., Foulley J.L., Chevalet C., Ollivier L., Genetic diversity of eleven European pig breeds, *Genet. Sel. Evol.* 32 (2000) 187–203.
- [26] Lebart L., Morineau A., Warwick K., *Multivariate Descriptive Statistical Analysis*, J. Wiley, New York, 1984.
- [27] Loftus R.T., MacHugh D.E., Ngere L.O., Balain D.S., Badi A.M., Bradley D.G., Cunningham E.P., Mitochondrial genetic variation in European, African and Indian cattle populations, *Anim. Genet.* 25 (1994) 265–271.
- [28] MacHugh D.E., Shriver M.D., Loftus R.T., Cunningham P., Bradley D.G., Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*), *Genetics* 146 (1997) 1071–86.
- [29] MacHugh D.E., Loftus R.T., Cunningham P., Bradley D.G., Genetic structure of seven European cattle breeds assessed using 20 microsatellite markers, *Anim. Genet.* 29 (1998) 333–40.
- [30] Mahé M.F., Miranda G., Queval R., Bado A., Souvenir Zafindrajaona P., Grosclaude F., Genetic polymorphism of milk proteins in African *Bos taurus* and *Bos indicus* populations. Characterization of variants α_{s1} -Cn H and β -Cn J, *Genet. Sel. Evol.* 31 (1999) 239–253.
- [31] Manwell C., Baker A.C.M., Biochemical classification of cattle. 2. Phylogenetic tree and specific status of the zebu, *Anim. Blood. Groups. Biochem. Genet.* 11 (1980) 151–162.
- [32] Martin-Burriel I., Garcia-Muro E., Zaragoza P., Genetic diversity analysis of six Spanish native cattle breeds using microsatellites, *Anim. Genet.* 30 (1999) 177–182.
- [33] Medjudorc I., Kustermann W., Lazar P., Russ I., Pirchner F., Marker-derived phylogeny of European cattle supports demic expansion of agriculture, *Anim. Genet.* 25 Suppl. 1 (1994) 19–27.
- [34] Minch E., Ruiz-Linares A., Goldstein D.B., Feldman M.W., Cavalli-Sforza L.L., MICROSAT (version 1.4): a computer program for calculating various statistics on microsatellite allele data, <http://lotka.stanford.edu:microsat.html>, 1995.

- [35] Moazami-Goudarzi K., Laloë D., Furet J.P., Grosclaude F., Analysis of genetic relationships between 10 cattle breeds with 17 microsatellites, *Anim. Genet.* 28 (1997) 338–345.
- [36] Mommens G., Peelman L.J., Van zeveren A., D'Ieteren G., Wissocq N., Microsatellite variation between an African and five European taurine breeds results in a geographical phylogenetic tree with a bison outgroup, *J. Anim. Breed. Genet.* 116 (1999) 325–330.
- [37] Nijman I.J., Bradley D.G., Hanotte O., Otsen M., Lenstra J.A., Satellite DNA polymorphisms and AFLP correlate with *Bos indicus-taurus* hybridization, *Anim. Genet.* 30 (1999) 265–273.
- [38] Queval R., Moazami-Goudarzi K., Laloë D., Mériaux J.C., Grosclaude F., Relations génétiques entre populations de taurins ou zébus d'Afrique de l'Ouest et taurins européens, *Genet. Sel. Evol.* 30 (1998) 367–383.
- [39] Reynolds J., Weir B.S., Cockerham C.C., Estimation of the coancestry coefficient: Basis for a short-term genetic distance, *Genetics* 105 (1983) 767–79.
- [40] Rohlf F.J., NTSYS-pc: Numerical Taxonomy and Multivariate Analysis System, Exeter Software, Heritage Lane, Setauket, New York, 1988.
- [41] Sambrook J., Fritsch E.F., Maniatis T., *Molecular Cloning: a Laboratory Manual*, 2nd ed., Cold Spring Harbor, Cold Spring Harbor Laboratory Press, NY, 1989.
- [42] Schmid M., Saitbekova N., Gaillard C., Dolf G., Genetic diversity in Swiss cattle breeds, *J. Anim. Breed. Genet.* 116 (1999) 1–8.
- [43] Schneider S., Kueffer J.M., Roessli D., Excoffier L., Arlequin ver 1.1: A software for population genetic data analysis, Genetics and Biometry Laboratory, University of Geneva, 1997.
- [44] Shriver M.D., Smith M.W., Jin L., Marcini A., Akey J.M., Dekas R., Ferrell R.E., Ethnic-Affiliation Estimation by use of population-specific DNA markers, *Am. J. Hum. Genet.* 60 (1997) 957–964.
- [45] Souvenir Zafindrajaona P., Zeuh V., Moazami-Goudarzi K., Laloë D., Bourzat D., Idriss A., Grosclaude F., Étude du statut phylogénétique du bovin Kouri du lac Tchad à l'aide de marqueurs moléculaires, *Rev. Elev. Méd. Vét. Pays Trop.* 52 (1999) 155–162.
- [46] Steffen P., Eggen A., Dietz A.B., Womack J.E., Stranzinger G., Fries R., Isolation and mapping of polymorphic microsatellites in cattle, *Anim. Genet.* 24 (1993) 121–124.
- [47] Swofford D.L., Selander R.B., BIOSYS-1. A computer program for the analysis of allelic variation in population genetics and biochemical systematics (Release 1.7), University of Illinois, Urbana, Champaign, 1989.
- [48] Takezaki N., Nei M., Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA, *Genetics* 144 (1996) 389–399.
- [49] Thaon d'Arnoldi C., Foulley J.L., Ollivier L., An overview of the Weitzman approach to diversity, *Genet. Sel. Evol.* 30 (1998) 149–161.
- [50] Vaiman D., Osta R., Mercier D., Grohs C., Leveziel H., Characterization of five new bovine dinucleotide repeats, *Anim. Genet.* 23 (1992) 537–541.
- [51] Vaiman D., Mercier D., Moazami-Goudarzi K., Eggen A., Ciampolini R., Lépling A., Velmala R., Kaukinen J., Varvio S.L., Martin P., Levéziel H., Guérin G., A set of 99 cattle microsatellites: characterization, synteny mapping, and polymorphism, *Mamm. Genome* 5 (1994) 288–297.

- [52] Weir B.S., Genetic Data Analysis, Sinauer Associates, Sunderland, Massachusetts, 1990.
- [53] Weitzman M., On diversity, *Quart. J. Econ.* 107 (1992) 363–405.
- [54] Weitzman M., What to preserve? An application of diversity theory to crane conservation, *Quart. J. Econ.* 108 (1993) 157–183.
- [55] Wright S., The interpretation of population structure by F-statistics with special regard to systems of mating, *Evolution* 19 (1965) 395–420.

Combining Inter- and Intrapopulation Information with the Weitzman Approach to Diversity Conservation

D. GARCÍA, N. CORRAL, AND J. CAÑÓN

From the Laboratorio de Genética, Facultad de Veterinaria, Universidad Complutense de Madrid, Spain (García and Cañón); and Departamento de Estadística, Investigación Operativa y D.M., Universidad de Oviedo, Spain (Corral).

Address correspondence to Javier Cañón, Laboratorio de Genética, Facultad de Veterinaria, Universidad Complutense de Madrid, Avda. Puerta de Hierro s/n, 28040—Madrid, Spain, or e-mail: jcanon@vet.ucm.es.

Abstract

This article introduces a new perspective on Weitzman's methodology for assessing the distribution of resources in genetic diversity conservation programs. Intrapopulation information is added to the procedure by diffusion process formulas to calculate genetic extinction probabilities, and therefore the marginal diversities and elasticities of diversity. The method was tested with a set of European cattle breeds from Spain and France and provided satisfactory results.

The efficient distribution of economic resources in biodiversity conservation programs is difficult since many economic, productive, morphological, genetic, social, and even affective and aesthetic factors are involved in decision making. In recent years, the preservation of biological diversity has been a headline topic of discussion [see Amos and Balmford (2001) for a short review]. In particular, the conservation of genetic diversity in livestock has received a great deal of attention (Barker 1999; Eding and Meuwissen 2001; Oldenbroek 1999; Ruane 2000). When developing a program for the conservation of genetic resources, special attention must be paid to how genetic diversity is measured, as it is well known that the rate of evolution in natural populations is limited by the degree of genetic variability (Fisher 1930). For small and endangered populations, it is important to avoid homozygosity as much as possible, because any increase leads to less genetic variation, and molecular estimates of inbreeding would suffer a parallel increase. This determines not only survival, but also adaptation to changing environments, including changes in consumer preferences and demands for animal-derived products (Frankham 1995; Hedrick and Kalinowski 2000; Lacy 1997). Considerable effort has been expended to minimize the long-term rate of inbreeding (Sonesson and Meuwissen 2001; Wang 1997; Wang and Hill 2000). Another point of particular interest is how to make decisions for investment policies to be applied to the different breeds or populations to maximize benefits in terms of diversity conservation.

Weitzman (1992) proposed a set of properties to be verified for the proper measurement of genetic diversity (see also Eding and Meuwissen 2001; Thaon d'Arnoldi et al. 1998). These properties make sense both intuitively and algebraically, but none of the usual measures of diversity verify them (Weitzman 1992). However, Weitzman developed a diversity function in which they all hold. This is ultimately based on pairwise distances between operational taxonomical units (OTUs), and it is easily implementable through a recursive algorithm. Since this is indirectly of the clustering type, it provides a graphic representation that has (in the strict sense) erroneously been called a "maximum-likelihood evolutionary tree." However, it supplies much useful information. Examples of its use can be found in Cañón et al. (2001), Laval et al. (2000), and Thaon d'Arnoldi et al. (1998). With this procedure, the contribution of each OTU, or groups of OTUs, to total diversity can easily be calculated, allowing conclusions on conservation policies to be drawn. However, as with any other methodology, there are some drawbacks (Caballero and Toro 2002; Eding and Meuwissen 2001). The most frequently mentioned is that it ignores within-population information, and being that assertion is not entirely true, since the method can be applied at any level—species, breed, population, and even individual—and therefore a within-population diversity can be computed, the fact is that the algorithm is a between-units one. A way of incorporating within-population information into the algorithm itself, involving computing Weitzman diversity within each population under study, is currently under

development. It is also commonly argued that it does not take into account different population sizes. The present work has filled these gaps by studying the development of gene frequencies within populations in order to compute extinction probabilities.

Weitzman's Approach

Weitzman (1993) devised a natural way of studying the evolution of diversity over different generations and of weighing the contributions of each OTU more precisely in order to assess the distribution of resources for conservation of genetic diversity programs. Basically it consists of calculating the expected value of a certain diversity measure—the method does not depend on which is chosen—by assigning probabilities to all the possible values of diversity at a particular time. Obviously there are as many values as subsets of the whole set of OTUs, since at a particular time (t) some OTUs may have become extinct and only the remainders count toward the final diversity. These values of probability depend on the probabilities of extinction of each particular OTU. Values are then obtained that measure how much the expected diversity at time t is affected by changes in each probability of extinction. Therefore a very logical and coherent, but conceptually simple way of weighing the OTUs with respect to their influence on expected diversity is constructed. This method has already been applied by Reist-Marti et al. (2003), but they do not escape the fact, already noted in Weitzman (1993), that the choice of these extinction probabilities is strongly subjective, and that is an important drawback of this approach. To circumvent this problem, the present investigation substitutes the concept of extinction with that of genetic extinction, in such a way that these probabilities can be calculated for any generation by using allelic frequencies and the theory of diffusion processes. In addition, this allows within-population diversity information to be brought into the analysis. The original procedure lacked this feature.

This method is a powerful tool for performing combined analyses integrating between- and within-population information. The former comes from pairwise distances, which can be genetic, and are studied by Weitzman's diversity measure. The latter come from allelic frequencies, and are studied in a second step using Weitzman's marginal diversity and elasticity of diversity (Weitzman 1993). These concepts are first introduced and later the theory of diffusion processes is applied to show how it accommodates Weitzman's model. Finally, an example of its use with a dataset from eight European cattle breeds is shown. All calculations were made using a FORTRAN program that implements the computation of Weitzman's diversity, including all the links, representatives, and percentages of diversity contributed by each OTU and each node of the tree. It also provides the "as if" ultra metric distance matrix, which, when used as an input in any tree-generating software, gives the unique hierarchical tree resulting from Weitzman's algorithm. Finally, for a given generation, the probabilities of genetic extinction, the marginal diversity, and the elasticity of the diversity of each OTU are derived.

Expected Diversity, Marginal Diversity, and Elasticity of Diversity

When establishing a policy of conservation of genetic resources, it is important not only to be able to evaluate current diversity, but to broaden the time horizon and infer the behavior of that diversity a number of generations ahead, and to determine which breeds or species are more influential upon it.

Assuming a trustworthy measure of diversity already exists—whether it is Weitzman's or another—it is of interest to know how the diversity of a set of populations will change, and which of them are more important for the conservation of genetic diversity in the context of biodiversity management programs. It is with this in mind that the concepts of expected diversity, marginal diversity, and elasticity of diversity are proposed.

Let \mathcal{Q} be a set of N_p OTUs. Let also $D: \wp(\mathcal{Q}) \rightarrow \mathbb{R}$ be a certain diversity function, and $\{P_i\}_{i=1, \dots, N_p}$ be the probabilities of extinction in one generation for each of them. The theoretical development of the procedure that will be described does not depend on D . It is applicable to any diversity function, as long as for any subset $S \subseteq \mathcal{Q}$, the value of $D(S)$ exists and is well defined. To exemplify the use of this methodology, the Weitzman diversity measure has been chosen. Its explicit definition will not be given here, since it can already be found and discussed in Weitzman (1992) and Thaon d'Arnoldi et al. (1998). If, for each subset of OTUs, $S \subseteq \mathcal{Q}$, the probability $P_t(S)$ of the OTUs in S having survived after t generations and those in $\mathcal{Q} \setminus S$ having become extinct can be found, then D at time t , D_t , can be thought of as a random variable whose expected value can easily be calculated. The expected diversity after t generations can therefore be defined as

$$ED_t = \sum_{S \in \wp(\mathcal{Q})} P_t(S) D_t(S). \quad (1)$$

This is not exactly the expression provided by Weitzman, since he defines the expected diversity as the cumulative sum of generations 1 to t , weighted by a "discount factor." The discount factor has been ignored here because, as shown below, the expected diversity is to be partially differentiated with respect to the probabilities of extinction. Therefore the discount factor provides no information for comparative purposes, and no advantage is to be found in accumulating the sum—besides, equation 1 is more coherent with the statistical concept of expected value.

Another important point that arises from the choice of Weitzman's diversity as the one to operate with is its very nature as a measure that uses pairwise distances as the input to calculate its value. If these distances do not vary with time, then $D_t(S) = D(S)$, the diversity at the present time. But $D_t(S)$ might be different from $D(S)$ if there is variation in the distances with time, as happens with most genetic distances. It is important therefore to assess the choice of the distance used to calculate the diversity, so that the effect of the chosen evolutionary model can be taken into account.

To calculate equation 1, it is assumed (Weitzman 1993) that the “lifetime” of population i follows an exponential distribution with parameter P_i . Thus $P_i(S)$ is given by

$$P_i(S) = \prod_{i \in S} \exp(-P_i t) \prod_{i \in Q \setminus S} (1 - \exp(-P_i t)). \quad (2)$$

Observe that ED_t is a function of (P_1, \dots, P_{N_P}) , and therefore can be partially differentiated with respect to each P_i . As a result, the marginal diversity for population i , $i = 1, \dots, N_P$, in generation t , is defined as

$$MD_t(i) = - \frac{\partial ED_t}{\partial P_i}(P_i). \quad (3)$$

With respect to the meaning of these expressions, the situation is as follows: a set of species, populations, or OTUs in general are under consideration; a measure of diversity has already been established for them; and the importance of each to conservation purposes can now be discussed. The future needs to be examined to see how that diversity might behave and how modifications to each population might affect that expected behavior. One of two things can happen to every single population: it may survive or it may become extinct. Therefore the diversity value at any future point in time will be that of the set of populations that have survived. Given the one-generation probabilities of extinction $\{P_i\}_{i=1, \dots, N_P}$, the probability of survival at time t of any possible subset of the original can easily be calculated, as already shown (equation 2). Now, for a set Q of N_P populations, there are 2^{N_P} possible subsets of Q , that is, there are 2^{N_P} possible patterns of survival/extinction, and hence 2^{N_P} different values of diversity. Each of these values can be calculated along with its corresponding probability, so it makes sense to consider diversity at time t as a random variable and obtain its expected value. This is precisely the expected diversity.

As just shown, this expected diversity is ultimately a function of $\{P_i\}_{i=1, \dots, N_P}$ so studies can be made of how variations in these probabilities affect the expected value of diversity. The more the diversity is affected by changes in a population's probability of extinction, the more that population should be considered a priority for action. But the concept of variation in this case is just that of a derivative, or more precisely, that of a partial derivative, with respect to each population's probability of extinction. Note that the value of these derivatives must be negative, since the more a population is likely to become extinct, the lower the expected diversity value will be. The marginal diversity is therefore defined as the negative of the value of the partial derivative to handle positive numbers.

Since a reduction in the probability of extinction provokes an increase in the expected diversity, and this increase is greater for higher marginal diversity values, it is on those populations with the highest values that attention should be focused. This must be directed toward lowering their extinction probability.

Another interesting indicator is the elasticity of diversity or the conservation potential of the population i , $i = 1, \dots, N_P$, in generation t . Elasticity is a very popular and well-known concept in economy. In most cases it compares the relative

variation in the quantity of demand with respect to the relative variation in prices, thereby obtaining a measure of the responsiveness, or sensitivity, of the demand to changes in price. If, for example, the price of a particular good increases by 2% and the demand for that good decreases by 5%, then the elasticity of the demand is -2.5 . Usually the elasticity of demand is negative, since an increase in prices has, as a consequence, a decrease in the quantity of demand. If the elasticity is less than -1 , then the demand is said to be elastic with respect to prices, while if it is between -1 and 0 , the demand is called inelastic. If the elasticity equals zero, the demand is completely inelastic, which means that the seller can raise the price of the product as much as he wants and the demand will not decrease at all. The other extreme is a perfectly elastic demand. In this situation, elasticity approaches $-\infty$, and the seller can sell virtually as much as he wants at the current market price. Raising the price would imply that demand would nearly disappear, and lowering it would be unprofitable, since demand is practically unlimited at the current price. Most real-life situations are, however, in the middle of these extremes, and sellers have to have adequate prices in order to optimize benefits.

The general formula for the price elasticity of demand is $E = ((\Delta Q_D / Q_D) / (\Delta P / P))$, where P is the current price of a certain good, Q_D is the quantity demanded at that price, ΔP is a small change in the current price, and ΔQ_D a small change in the quantity demanded. Note that this expression can be rearranged as $E = (\Delta Q_D / \Delta P) (P / Q_D)$. When the changes are not measured in a short time period, averages of the initial and final values are taken on Q_D and P . To approximate the elasticity at a particular point, short intervals can be taken around quantity and price values at that point to calculate differences. However, when differentiation is possible, exact computation is available, as will be shown in equation 4.

The concept of elasticity can be applied to many other economic variables, such as supply instead of demand, or in noneconomic contexts, as happens in this case. An analogy can be established here between quantity of demand and expected diversity value and between price and extinction probability, so the elasticity of diversity or conservation potential, as renamed in Weitzman (1993), can be defined as

$$CP_t(i) = - \frac{\partial ED_t}{\partial P_i}(P_i) \frac{P_i}{ED_t}, \quad (4)$$

where, as before, the minus sign is added to deal with positive numbers.

Elasticity is particularly relevant when the cost of reducing the probability of extinction of a population is directly proportional to the probability itself [see Weitzman (1993) for further discussion on the singularities of equations 3 and 4].

Genetic Extinction

All the previous definitions are based on one-period probabilities of extinction of each population—which ought to be known. However, this is not as easy as it sounds. There is no objective way of determining this for a common species, population, or breed. Therefore the information obtained

depends on the subjectiveness of the probability values, which is not very desirable. Scientific, rigorous, and objective probabilities are required if the above methodology is to be correctly applied. Instead of the concept of physical extinction, it is proposed here that genetic extinction be used. The concept is even more appropriate if it is taken into account that it is genetic diversity that is being referred to in a framework of genetic resources conservation programs. In addition, this introduces within-breed information from allelic frequencies into the analysis.

The concept of genetic extinction is closely related to that of homozygosity, since the genetic extinction of a population is equivalent to a homozygosity rate of one. The more homozygous a population is, the more genetically endangered. But homozygosity is, in practice, nothing more than an average of a number of homozygous individuals taken across a series of representative loci. Therefore, by analogy, an average can be calculated for the allelic situation of the loci under study for each breed. There are two alternatives. One is to average the probability of fixation of the loci at time t and use that to calculate the marginal diversity and conservation potential. The other is to perform the analysis on each locus with its corresponding fixation probability, obtain a different value of marginal diversity and conservation potential for each, and average those values across the loci. The extinction probabilities can be loosely extended to the point in which allelic frequencies exceed a certain value α instead of just equaling one.

For ease of notation, index i will be dropped in the following formulas, which is referred to a generic population, but all the calculations will be ultimately applied to each of the N_p populations to obtain the corresponding extinction probabilities.

Consider a set of L loci, and for locus l , let $\{A_1^l, \dots, A_{n_l}^l\}$ be the alleles, $l = 1, \dots, L$. For each $m = 1, \dots, n_l$ and $l = 1, \dots, L$, let $p_m^l(t)$ be the frequency of the allele A_m^l in generation t in a certain population. For the first option, $P_{ext}(t)$ is therefore calculated as

$$\begin{aligned} P_{ext}(t) &= \frac{1}{L} \sum_{l=1}^L P \left[\bigcup_{m=1}^{n_l} (p_m^l(t) \geq \alpha) \right] \\ &= \frac{1}{L} \sum_{l=1}^L \sum_{m=1}^{n_l} P[p_m^l(t) \geq \alpha], \end{aligned} \quad (5)$$

the second equality holding because the events are disjointed (if one allele becomes fixed or has a frequency higher than α , for $\alpha > 0.5$, no other allele can do the same). For the second option, a set of L probabilities $\{P_{ext}(t, l)\}_{l=1, \dots, L}$ can be calculated:

$$\begin{aligned} P_{ext}(t, l) &= P \left[\bigcup_{m=1}^{n_l} (p_m^l(t) \geq \alpha) \right] \\ &= \sum_{m=1}^{n_l} P[p_m^l(t) \geq \alpha], \quad l = 1, \dots, L. \end{aligned} \quad (6)$$

The change in gene frequencies over the generations is an intricate process that depends on very different factors.

However, diffusion processes allow explicit expressions for $P[p_m^l(t) \geq \alpha]$ in many cases. To illustrate the approach introduced in this article, the particular situation of populations solely under the effect of genetic drift will be explained. In this case, it can be proven that

$$\begin{aligned} P_{ext}(t) &= \frac{1}{L} \sum_{l=1}^L \sum_{m=1}^{n_l} \left[p_m^l(0) + \sum_{i=1}^{\infty} (2i+1) p_m^l(0) \right. \\ &\quad \times (1 - p_m^l(0)) F(i+2, 1-i, 2, p_m^l(0)) \\ &\quad \times [F(-i, i+1, 1, \alpha) - F(-i, i+1, 1, 1)] \\ &\quad \left. + (-1)^i \right] e^{-\frac{i(i+1)t}{4N}} \end{aligned} \quad (7)$$

and

$$\begin{aligned} \{P_{ext}(t, l)\}_{l=1, \dots, L} &= \left\{ \sum_{m=1}^{n_l} \left[p_m^l(0) + \sum_{i=1}^{\infty} (2i+1) p_m^l(0) \right. \right. \\ &\quad \times (1 - p_m^l(0)) F(i+2, 1-i, 2, p_m^l(0)) \\ &\quad \times [F(-i, i+1, 1, \alpha) - F(-i, i+1, 1, 1)] \\ &\quad \left. \left. + (-1)^i \right] e^{-\frac{i(i+1)t}{4N}} \right\}_{l=1, \dots, L}, \end{aligned} \quad (8)$$

where $\{p_m^l(0)\}_{m=1, \dots, n_l}^{l=1, \dots, L}$ are the present-time allelic frequencies of loci 1 to L , and $F(\cdot, \cdot, \cdot, \cdot)$ is the hypergeometric function.

Equations 7 and 8 are calculated for a single population and require the allelic frequencies of all the marker loci involved in the study. For a whole set of populations, a set of probabilities exists, $\{P_{ext}^i(t)\}_{i=1, \dots, N_p}$ (or $\{P_{ext}^i(t, l)\}_{i=1, \dots, N_p}^{l=1, \dots, L}$), each of them calculated according to equation 7 or 8. Observe that these probabilities are not conceptually the same as those in $\{P_i\}_{i=1, \dots, N_p}$, since the latter are one-generation extinction probabilities, while the former are the direct probabilities of extinction at generation t .

Equations 7 and 8 allow that which is desired: to have at our disposal a series of objective probability values based on genetic data. Furthermore, with this approach, within-population information as well as information on population sizes is introduced into the analysis.

For ease of notation, only the first case, that of averaged probability values across loci will be analyzed, so probabilities, $\{P_{ext}^i(t)\}_{i=1, \dots, N_p}$, obtained for each population according to equation 7, will be used. Thus, for a subset S of the whole set of populations \mathcal{Q} , equation 2 is substituted by

$$P_t(S) = \prod_{i \in S} (1 - P_{ext}^i(t)) \prod_{i \in \mathcal{Q} \setminus S} P_{ext}^i(t) \quad (9)$$

in equation 1 to calculate the expected diversity. The definitions of marginal diversity and elasticity would be analogously modified, so the differentials would be made with respect to $P_{ext}^i(t)$ instead of P_i .

Note that with this model, ED_t is a linear function of each $P_{ext}^i(t)$ separately. This is, if for a given i , we consider

$P_{ext}^i(t)$ as variable and the rest, $P_{ext}^j(t)$, $i \neq j \in \{1, \dots, N_p\}$ as fixed, then

$$\begin{aligned}
 ED_t(i) &= \sum_{S \in \wp(Q)} D(S) P_t(S) \\
 &= \sum_{S \in \wp(Q)} D(S) \prod_{j \in S} (1 - P_{ext}^j(t)) \prod_{j \in Q \setminus S} P_{ext}^j(t) \\
 &= \begin{cases} \sum_{S \in \wp(Q)} D(S) (1 - P_{ext}^i(t)) \prod_{j \in S, j \neq i} (1 - P_{ext}^j(t)) \\ \quad \times \prod_{j \in Q \setminus S, j \neq i} P_{ext}^j(t), & \text{if } i \in S \\ \sum_{S \in \wp(Q)} D(S) P_{ext}^i(t) \prod_{j \in S, j \neq i} (1 - P_{ext}^j(t)) \prod_{j \in Q \setminus S, j \neq i} P_{ext}^j(t), & \text{if } i \notin S \end{cases}
 \end{aligned} \quad (10)$$

Therefore $MD_t(i)$, being the partial differential of ED_t with respect to $P_{ext}^i(t)$, is a constant value for each $i = 1, \dots, N$. In fact,

$$\begin{aligned}
 MD_t(i) &= \begin{cases} \sum_{S \in \wp(Q)} \frac{\partial}{\partial P_{ext}^i(t)} D(S) (1 - P_{ext}^i(t)) \prod_{j \in S, j \neq i} (1 - P_{ext}^j(t)) \\ \quad \times \prod_{j \in Q \setminus S, j \neq i} P_{ext}^j(t), & \text{if } i \in S \\ \sum_{S \in \wp(Q)} \frac{\partial}{\partial P_{ext}^i(t)} D(S) P_{ext}^i(t) \prod_{j \in S, j \neq i} (1 - P_{ext}^j(t)) \\ \quad \times \prod_{j \in Q \setminus S, j \neq i} P_{ext}^j(t), & \text{if } i \notin S \end{cases} \\
 &= \begin{cases} \sum_{S \in \wp(Q)} -D(S) \prod_{j \in S, j \neq i} (1 - P_{ext}^j(t)) \prod_{j \in Q \setminus S, j \neq i} P_{ext}^j(t), & \text{if } i \in S \\ \sum_{S \in \wp(Q)} D(S) \prod_{j \in S, j \neq i} (1 - P_{ext}^j(t)) \prod_{j \in Q \setminus S, j \neq i} P_{ext}^j(t), & \text{if } i \notin S \end{cases}
 \end{aligned} \quad (11)$$

so therefore

$$MD_t(i) = - \sum_{S \in \wp(Q)} D(S) P_t^*(S \setminus i) (-1)^{I_S(i)}, \quad (12)$$

where $I_S(i)$ equals one when $i \in S$ and zero when $i \notin S$, and P_t^* is defined over $\wp(Q \setminus i)$ in the same way that P_t is over $\wp(Q)$, and consequently $MD_t(i)$ does not actually depend on $P_{ext}^i(t)$. Had one-period extinction probabilities been employed, such as in Weitzman (1993), there would be a non-linear dependence of ED_t on each $P_{ext}^i(t)$, and $MD_t(i)$ would indeed depend on $P_{ext}^i(t)$. Now, from equation 12, $MD_t(i) = \sum_{S \in \wp(Q)} D(S) P_t^*(S \setminus i) - \sum_{S \in \wp(Q)} D(S) P_t^*(S)$, and, finally, equation 12 can be rewritten as

$$MD_t(i) = - \sum_{S \in \wp(Q \setminus i)} P_t^*(S) [D(S \cup i) - D(S)]. \quad (13)$$

Table 1. Breeds, countries of origin, and effective sizes

Breed	Country	Effective size
Alistana (ALI)	Spain	36
Asturiana de Montaña (ASM)	Spain	35
Asturiana de Valles (ASV)	Spain	89
Aubrac (AUB)	France	100
Gasconne (GAS)	France	100
Salers (SAL)	France	100
Sayaguesa (SAY)	Spain	21
Tudanca (TUD)	Spain	35

The expression in brackets recalls strongly the partial contribution of population i , calculated for every $i = 1, \dots, N_p$ as $PC_i = [D(Q) - D(Q \setminus i)]/D(Q)$. In fact, if a random variable $C_i: \wp(Q \setminus i) \rightarrow \mathbb{R}$ is defined such that for every $S \subseteq Q \setminus i$, $C_i(S) := D(S \cup i) - D(S)$ with probability $P_t^*(S)$, then the marginal diversity is the expected value of C_i at generation t , and, divided by the expected diversity at generation t , can be understood as an expected partial contribution at generation t .

Application and Results

This procedure was tested with an example set of local French and Spanish cattle breeds from the European contract FAIR1 CT95 0702 project [see Cañón et al. (2001) for details]. Table 1 shows the names, origins, and effective sizes of the breeds.

Allelic frequencies from 16 microsatellite-type marker loci, obtained from a total of 50 animals per breed (Cañón et al. 2001), were used to calculate marginal diversities and elasticities. Goldstein et al. (1995) proposed a distance measure, the average squared distance, especially devised for markers with a high degree of polymorphism, such as microsatellites. Some of its properties include the fact that its expected value does not change over time under a genetic drift model, which allows for $D_t(S) = D(S)$. Although it is not a distance in the strict sense, since when measured from one population to itself it does not equal zero, the procedure can be applied without loss of generality. Thus this distance was applied to calculate the Weitzman diversity, and partial contributions of the breeds to the total current diversity, $\{PC_i\}_{i=1, \dots, N_p}$, were obtained. Note that this distance is used just as an example, since the method presented here can be implemented with any distance, as long as its variation with time can be assessed effectively enough. Actually the distance measure does not necessarily have to be a genetic distance. It may involve morphological or other kinds of information, and it is up to the person in charge of putting the method into practice to decide which particular distance suits the situation under study.

Table 2 shows the partial contributions and the marginal diversities and conservation potentials obtained for 25, 35, and 50 generations, and an α threshold of 0.90. Values were rescaled to 100. Other within-population variation statistics

Table 2. Mean expected (MEH) and observed (MOH) homozygosities, mean effective number of alleles (MENA), partial contributions (PC), and average probabilities of extinction ($P_{\text{ext}}^i(t)$), marginal diversities ($MD_i(t)$), and conservation potentials ($CP_i(t)$) from averaged probabilities for $t = 25, 35$, and 50 , using the average squared distance

Breed	MEH	MOH	MENA	$P_{\text{ext}}^i(25)$	$P_{\text{ext}}^i(35)$	$P_{\text{ext}}^i(50)$	PC	$MD_i(25)$	$MD_i(35)$	$MD_i(50)$	$CP_i(25)$	$CP_i(35)$	$CP_i(50)$
SAL	0.369	0.420	4.35	0.030	0.050	0.084	23.778	23.830	23.846	23.817	15.038	14.182	14.100
AUB	0.389	0.431	4.68	0.029	0.054	0.093	19.009	19.054	19.088	19.123	11.723	12.208	12.532
GAS	0.292	0.370	3.89	0.005	0.014	0.034	25.686	25.673	25.652	25.565	2.780	4.256	6.154
ALI	0.319	0.371	4.01	0.079	0.151	0.267	5.817	5.818	5.827	5.856	9.653	10.469	11.025
ASM	0.295	0.348	4.53	0.076	0.146	0.263	6.634	6.566	6.528	6.495	10.522	11.351	12.028
ASV	0.317	0.344	6.49	0.025	0.043	0.075	5.025	5.051	5.087	5.206	2.630	2.615	2.755
SAY	0.293	0.346	3.88	0.194	0.326	0.504	8.184	8.165	8.149	8.128	33.137	31.618	28.853
TUD	0.349	0.404	3.74	0.119	0.192	0.306	5.868	5.842	5.823	5.809	14.517	13.301	12.553

are included, such as mean observed and expected homozygosity and mean effective number of alleles. The results from the two methods described above—average extinction probabilities and average marginal diversities and conservation potentials—are extremely similar (data not shown). Suffice it to say, the mean coefficient of variation of the marginal diversities and the conservation potentials were 0.004% and 0.09%, respectively. Therefore, for the sake of simplicity, only values for the first method (average extinction probability across loci) are shown. Table 2 also shows that the differences in the figures among the breeds for other traditional within-variation measures are too small to make clear distinctions.

According to the partial contributions, Gasconne and Salers are the two most contributing breeds, followed by Aubrac, and, by far, by the Sayaguesa breed. Alistana and Asturiana de Valles are the breeds with the lowest partial contributions.

Marginal diversities are almost identical to partial contributions. The ultimate reason is the very nature of the marginal diversity. As seen in equation 13, marginal diversity is a sort of expected partial contribution at the given generation. Now, in the weighted sum, the set $Q \setminus i \in \wp(Q \setminus i)$ has a higher weight, since the probability of survival of all the populations in $Q \setminus i$ is noticeably higher than the probability of any other combination of survival/extinction within the set of populations $Q \setminus i$. This happens especially in short to medium time intervals, because the fewer the number of generations, the lower the fixation probability for any allele and therefore the lower the extinction probability, making the product of the probabilities used to calculate each term of equation 13—except for $Q \setminus i$ —much smaller. This fact may lead one to consider marginal diversity as a superfluous indicator once we already have partial contributions, but it may be seen so only in the short term, because as generations pass, the differences between partial contributions and marginal diversities increase. This increase is already suggested in Table 2, although it would be much more evident for a larger number of generations.

Conservation potentials offer a different prioritization scheme than marginal diversities. Because of its low effective size, Sayaguesa is now the highest scored breed, followed by Salers, due to its strong influence in the between-population variation. Salers and Aubrac, highly prioritized with the mar-

ginal diversity criterion, are also important with respect to the conservation potential. Excluding Sayaguesa, Salers and Aubrac would join Alistana, Asturiana de Montaña, and Tudanca in a middle group. Differences within this group tend to diminish notably as the time horizon increases. The figures for Gasconne are particularly noteworthy. Marginal diversity sets this breed as the first in the prioritization ranking, while its conservation potentials are remarkably low—only Asturiana de Valles shows lower values. This may possibly be due to the fact that Gasconne is uniformly the most distant breed, that is, for any certain breed, Gasconne has a higher pairwise distance with this breed than any other one (data not shown), so it scores high in partial contribution and marginal diversity, where between-breed information seems to have a greater weight. On the other hand, it has the lowest extinction probability of all the breeds (down to six times that of Salers and Aubrac; all three having the same effective size), and this pulls down its ranking in the conservation potential. Only Asturiana de Valles is below Gasconne. This might be due to the combination of having a relatively low extinction probability and being, as opposed to Gasconne, uniformly the closest breed, meaning that for any certain breed, Asturiana de Valles has a lower pairwise distance with this breed than any other one (data not shown). So, on one side, within-breed information says that its extinction probability is low, while on the other, between-breed information says that it is quite close to the rest of the breeds. Therefore it has the lowest conservation potential value.

Discussion

Maintaining genetic variability is an important goal in the conservation of animal populations and Weitzman's procedure can be applied to achieve this. However, it has been criticized (Caballero and Toro 2002; Eding and Meuwissen 2001; Thaon d'Arnoldi et al. 1998) for not considering within-population variation. Only Ollivier and Foulley (2002) suggest the otherwise obvious, but always ignored, possibility of applying Weitzman's algorithm to obtain a within-diversity measure. To do this, the coefficient of molecular coancestry (Fabel et al. 2004) between individuals, which applies Malécot's (1948) definition, but uses the concept of identity

in state instead of identity by descent, could be used to build a matrix of genetic distances (actually $1 - \text{Malécot's coefficient}$) between individuals within one population to compute the Weitzman diversity measure within each population. However, this way of calculating within-breed diversity is actually highly correlated ($r = -0.985$ in the example presented in this article; data not shown) with within-breed heterozygosity. Other criteria have been proposed to estimate within-breed variation, such as allelic richness (Marshall and Brown 1975). Efficiency of heterozygosity or allelic richness depends on the interest in the short- or long-term selection response.

If no extinction probabilities, be they genetic based or not, and no future behavior of the diversity and each population's contributions to it are to be studied, other solutions to incorporate Weitzman's approach to a joint between- and within-population analysis have been suggested. Ollivier and Foulley (2002) proposed a linear combination of between-population diversity, calculated via the Weitzman approach, and within-population diversity, measured with the heterozygosity, as an aggregate diversity, similar in concept to the traditional aggregate genotype in animal breeding. Depending on the emphasis given to the different possible breeding objectives, for example, selection or crossbreeding, different weights can be justified. It can be shown (Ollivier and Foulley 2004) that when the weights of between-breed and within-breed contributions are proportional to F_{ST} (between-population differentiation index) and $(1 - F_{ST})$, breeds rank in a similar way when applying the method proposed by Petit et al. (1998) and Caballero and Toro (2002).

It is well known that more than 80% of the total variability in a domestic animal species is a consequence of the genetic differences among individuals within subpopulations, (e.g., breeds), so giving a relatively greater emphasis to this within-breed component means promoting the possibility of the selection response within breeds. However, the crossbreeding, to exploit the heterosis or the complementarity, is also an important genetic improvement strategy to be used in animal production, and their magnitudes are proportional to the genetic distances among breeds (Falconer and Mackay 1996:255). This should justify giving more emphasis to the between-breed component, as illustrated by Chaiwong and Kinghorn (1999).

Using extinction probabilities allows a different way of using Weitzman's methodology. Between- and within-genetic information are both included in the analysis, and the behavior of the diversity in future generations can be assessed. The two alternate definitions of extinction probabilities are computable without difficulty and the rationale behind them agrees with that of popular measures like heterozygosity. The results from both procedures do not differ too much and allow clear differentiations among the breeds, something not possible with more classical measures of within-population variation, as can be deduced from Table 2. The close relationship between the rate of inbreeding and the probability of extinction defined here allows the application of Weitzman's methodology to classify populations by the effect on expected diversity when efforts (Sonesson and

Meuwissen 2001; Wang 1997; Wang and Hill 2000) are made to reduce their homozygosity increase per generation.

As discussed above, marginal diversities were similar to partial contributions in all breeds, which supports the concept of expected partial contributions. Nevertheless, there are upward trends in some of the breeds and downward trends in others, which reflects that although the most influential factor for the marginal diversity seems to be the weight of the breed in terms of diversity, extinction probabilities tend to force constant patterns of increase/decrease of the marginals. This perhaps excessive influence of the phylogenetic scene on marginal diversities might be overcome by introducing one-generation genetic extinction probabilities, such as in Weitzman's work (1993), and basing all the analysis at any generation on them. However, this is beyond the scope of this article.

Neither is it the aim of this article to decide or advise on which of the many available distance measures should be used to perform an analysis. On the contrary, the methodology presented here is introduced as a general procedure in which a distance measure is required, but its choice is not inherent to the procedure. However, the particulars of each individual application should be taken into account when choosing the distance to use, since different distances can actually lead to quite different conclusions. For example, since the effective size already has an important weight on the extinction probability, the more a distance weight drifts, the more the values for marginals and elasticities will correlate with drift-based measures such as heterozygosity. Different distances should be used depending on the kind of OTUs or taxas used. Some distances are better suited for species comparisons and others for close breed analysis, depending on the time since divergence and the drift-mutation model. The same applies to the solution of the Kolmogorov (1931) equation, used to obtain equations 7 and 8, which can be found for a number of different scenarios.

The example shows that the results for the partial contributions and marginal diversities in short- or medium-term applications are extremely similar. Since no within-population information is taken into account to obtain the partial contributions, it seems that the marginal diversity is not a very good indicator to account for both between- and within-population information in short- or medium-term applications. In the long term, the marginal diversity seems to depart from the value of partial contribution, so it should be interesting to consider it.

Regarding conservation potentials, figures show that effective size is an important factor, as proved by Sayaguesa's high values, but also that other important information is accounted for in the model, so the final result is a combination of all the genetic properties of the populations.

Also, as noted by Weitzman (1993), elasticity, or conservation potential, is the most trustworthy indicator when the mechanism of investment is directly related to extinction probability, since the marginal diversity is weighted by the extinction probability itself. Ideally an optimal study should include a functional relationship between costs and effects on extinction probability, so the problem could be examined in

terms of function optimization subject to a series of constraints. In the absence of this kind of information, a global view of the results must be taken to arrive at balanced advice. In addition, it must not be forgotten that many factors other than genetics are usually and necessarily taken into account when addressing the conservation of biological diversity, and the information provided by genetic diversity studies must not be considered the sole source of information for decision making. Nevertheless, the methodology introduced here gathers together different types of genetic data (between breed, within breed, and population size) and is a useful tool in diversity conservation assessment.

The kind of analysis presented is recommended when mid- or long-term diversity studies are proposed, since genes, on average, are not expected to suffer noticeable risk of fixation in short time periods, unless effective sizes are particularly small, in which case a lesser number of generations can be considered.

As a final consideration, the appendix details a new approximation of the Weitzman algorithm that allows for a greater number of populations to be introduced in the analysis and improves existing approximations.

Appendix

The Weitzman algorithm (Weitzman 1992) is very computer time demanding, and computation time grows exponentially with the number of populations included in the analysis—log regression of the experimental computing time versus the number of populations provides an r^2 of 0.998 (data not shown). For example, with an AMD processor at 2.4 GHz, computing only the diversity of 33 populations, without partial contributions or marginals, took almost 11 h, and time increases by an estimated factor of 2.14 (data not shown), so handling larger numbers of populations becomes infeasible. Therefore some kind of approximation of the exact algorithm must be used if the number of populations becomes larger than a few dozen.

Thaon d'Arnoldi et al. (1998) proposed an approximation consisting of randomly sampling trees among the 2^{n-1} that the algorithm examines and taking as the approximated value of diversity the maximum of the values coming out of the sampled trees. This procedure works well with a relatively small number of populations—they tested it up to 29—but as the number of populations increases, so must the sample size. The problem is that while the exact algorithm can be implemented recursively, the approximated one cannot, and it comes to a point where, for reasonable precision in the estimation, the approximated algorithm actually takes more time than the exact one.

We propose a different approximation, and preliminary results seem to indicate that its behavior is rather satisfactory. In each step, the algorithm splits into two symmetrical recursive problems, each of which deals with a submatrix one dimension smaller than the input one. What we propose is, before inputting these submatrices into the next step, checking their dispersion, and if all the values in one submatrix are

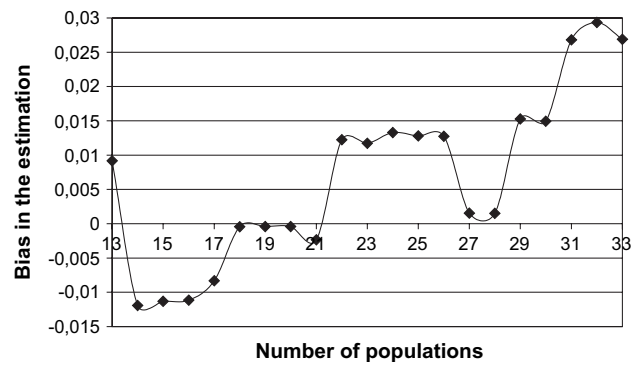


Figure 1. Plot of the bias between real diversity values and estimated ones, calculated as the estimated value minus the real value divided by the real value, for a limit of 0.25 in the coefficient of variation for the approximation.

close enough together around a mean value, to use this value, multiplied by the remaining number of steps, as the diversity value for that submatrix, and if not, then proceed with the exact computation in the usual way. The way to check for this “closeness” is to set a threshold to the variation coefficient (i.e., the standard deviation divided by the mean) of the values in the matrix. If this coefficient is lower than the threshold, then approximate, and if it is larger, then go on with the exact algorithm—which might eventually be truncated in some further step if the condition is met.

Figure 1 shows the bias we observed between the real and the approximated values of the diversity through increasing numbers of populations, from 13 to 33, and it can be seen that the bias never goes beyond 3% in absolute value. Calculations were made for a threshold on the variation coefficient of 0.25. Of course, the higher the threshold, the faster the computation, but with a higher bias attached, and vice versa. The behavior of the bias, as shown in Figure 1, is so irregular because the approximation is quite dependent on the data structure—the distance matrix in this case—but apparently it seems to provide good results. For 33 populations, the exact algorithm took, as mentioned above, almost 11 h to complete, while on the same computer the approximation took 2 min 40 s, with a bias of 2.7%. For a different dataset, 35 populations took 2 days minus 5 min until completion with the exact algorithm, and the approximation took less than 2 s with a threshold of 0.4, biasing the result less than 2%.

As promising as these results may seem, this procedure should be tested exhaustively before being applied to large datasets on the order of 50 or more populations, but that task is beyond the scope of this article.

Acknowledgments

We thank A. Burton for providing language advice and the reviewers for their helpful comments on this article. This study received the financial support of FEDER project 2FD1997-1191 and European project CT-98-118.

References

- Amos W and Balmford A, 2001. When does conservation genetics matter? *Heredity* 87:257–265.
- Barker JSF, 1999. Conservation of livestock breed diversity. *Anim Genet Resource Inform* 25:33–43.
- Caballero A and Toro M, 2002. Analysis of genetic diversity for the management of conserved subdivided populations. *Conserv Genet* 3:289–299.
- Cañón J, Alexandrino P, Bessa I, Carleos C, Carretero Y, Dunner S, Ferran N, García D, Jordana J, Laloë D, Pereira A, Sanchez A, and Moazami-Goudarzi K, 2001. Genetic diversity measures of local European beef cattle breeds for conservation purposes. *Genet Sel Evol* 33:311–332.
- Chaiwong N and Kinghorn BP, 1999. Use of genetic markers to aid conservation decisions for groups of rare domestic breeds. *Proc Assoc Adv Anim Breed Genet* 13:365–368.
- Eding H and Meuwissen T, 2001. Marker-based estimates of between and within population kinships for the conservation of genetic diversity. *J Anim Breed Genet* 118:141–159.
- Fabuel E, Barragán C, Silió L, Rodríguez MC, and Toro MA, 2004. Analysis of genetic diversity and conservation priorities in Iberian pigs based on microsatellite markers. *Heredity* 93:104–113.
- Falconer DS and Mackay TFC, 1996. An introduction to quantitative genetics, 4th ed. Harlow, UK: Longman.
- Fisher RA, 1930. The genetical theory of natural selection. Oxford: Clarendon Press.
- Frankham R, 1995. Inbreeding and conservation: a threshold effect. *Conserv Biol* 9:792–799.
- Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, and Feldman MW, 1995. An evaluation of genetic distances for use with microsatellite loci. *Genetics* 139:463–471.
- Hedrick PW and Kalinowski ST, 2000. Inbreeding depression in conservation biology. *Annu Rev Ecol Syst* 31:139–162.
- Kolmogorov A, 1931. Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. *Math Ann* 104:415–458.
- Lacy RC, 1997. Importance of genetic variation to the viability of mammalian populations. *J Mamm* 78:320–335.
- Laval G, Iannuccelli N, Legault C, Milan D, Groenen MAM, Giuffra, E, Andersson L, Nissen PH, Jorgensen CB, Beeckmann P, Geldermann H, Foulley JL, Chevalet C, and Ollivier L, 2000. Genetic diversity of eleven European pig breeds. *Genet Sel Evol* 32:187–203.
- Malécot G, 1948. Les Mathématiques de l'Hérédité. Paris: Masson et Cie.
- Marshall DR and Brown AHD, 1975. Optimum sampling strategies in genetic conservation. In: *Crop genetic resources for today and tomorrow* (Frankel OH and Hawkes JG, eds). Cambridge: Cambridge University Press; 53–80.
- Oldenbroek JK, 1999. Genebanks and the conservation of farm animal genetic resources. Lelystad, The Netherlands: DLO Institute for Animal Science.
- Ollivier L and Foulley JL, 2002. Some suggestions on how to preserve both within- and between-breed genetic diversity. In: *Book of Abstracts of the 53rd Annual Meeting of the European Association for Animal Production*, Cairo, Egypt, 1–4 September 2002 (van der Honing Y, ed). Wageningen, The Netherlands: Wageningen Academic Publishers.
- Ollivier L and Foulley JL, 2004. Objectives in livestock diversity preservation: the European pig example. In: *Wissenschaftliches Kolloquium "Nutztierzüchtung im Wandel der Zeit."* Göttingen, Germany: Cuvillier Verlag; 87–106.
- Petit RJ, El Mousadik A, and Pons O, 1998. Identifying populations for conservation on the basis of genetic markers. *Conserv Biol* 12:844–855.
- Reist-Marti SB, Gibson J, Rege JEO, Simianer H, and Hanotte O, 2003. Weitzman's approach and conservation of breed diversity: an application to African cattle breeds. *Conserv Biol* 17:1299–1311.
- Ruane J, 2000. A framework for prioritizing domestic animal breeds for conservation purposes at the national level: a Norwegian case study. *Conserv Biol* 14:1385–1393.
- Sonesson AK and Meuwissen THE, 2001. Minimization of rate of inbreeding for small populations with overlapping generations. *Genet Res* 77: 285–292.
- Thaon d'Arnoldi C, Foulley JL, and Ollivier L, 1998. An overview of the Weitzman approach to diversity. *Genet Sel Evol* 30:149–161.
- Wang J, 1997. Effective size and F-statistics of subdivided populations. 2. Dioecious species. *Genetics* 146:1465–1474.
- Wang J and Hill WG, 2000. Marker assisted selection to increase effective population size by reducing Mendelian segregation variance. *Genetics* 154:475–489.
- Weitzman ML, 1992. On diversity. *Q J Econ* 107:363–405.
- Weitzman M, 1993. What to preserve? An application of diversity theory to crane conservation. *Q J Econ* 108:157–183.

Received June 23, 2004

Accepted June 23, 2005

Corresponding Editor: Leif Andersson

Genetic variation within the Lidia bovine breed

J. Cañón*, I. Tupac-Yupanqui*, M. A. García-Atance*, O. Cortés*, D. García*, J. Fernández[†] and S. Dunner*

*Department of Animal Production, Facultad de Veterinaria, Universidad Complutense, 28040 Madrid, Spain. [†]Unión de Criadores de Toros de Lidia, Departamento Técnico, 28010 Madrid, Spain

Summary

The results of an exhaustive data collection from a bovine population with a low level of exchangeability, the Lidia breed, are presented. A total of 1683 individuals from 79 herds were sampled and genetic diversity within and among lineages was assessed using 24 microsatellite loci on 22 different chromosomes. Expected heterozygosity ranged between 0.46 and 0.68 per lineage and there was significant inbreeding in the lineages, which included several farms [mean F_{IS} = 0.11, bootstrap 95% confidence interval (0.09, 0.14)], mainly because of the high genetic divergence between herds within those lineages. High genetic differentiation between lineages was also found with a mean F_{ST} of 0.18 [bootstrap 95% confidence interval (0.17, 0.19)], and all pairwise values, which ranged from 0.07 to 0.35, were highly significant. The relationships among lineages showed weak statistical support. Nonetheless, lineages were highly discrete when analysed using correspondence analysis and a great proportion of the individuals were correctly assigned to their own lineage when performing standard assignment procedures.

Keywords bovine, fighting bull, genetic diversity, microsatellite.

Introduction

Most studies usually treat breeds as the management unit within the species, failing to analyse the within-breed structure (Buchanan *et al.* 1994; Blott *et al.* 1998), thus neglecting potential within-breed differentiation. However, the analysis of the genetic structure of certain breeds can also be of interest. Such might be the case of the Lidia bovine breed. The Lidia bovine breed, otherwise known as the fighting bull, is the most successful domestic breed of any Spanish animal, being one of the largest and most inclusive intergrading bovine breeds. It has spread through European countries, such as France, Portugal and Spain, and numerous Central and South American countries (<http://www.toroslidia.com/>).

This breed has some peculiarities, which make it a population that deserves separate genetic analysis. First, it is the only bovine population selected for behavioural traits with a long history of isolation from the rest of the domestic bovine

breeds (see Silva *et al.* 2002, 2006 for the registered behavioural traits). Second, the term Lidia refers to a racial grouping of native bovines of the Iberian Peninsula, which survive almost exclusively in the Mediterranean forest ecosystem traditionally known as La Dehesa (pastureland interspersed with Mediterranean oaks). Their behavioural characteristics, the uniqueness of their management and the hundreds of years of genetic isolation make it difficult to find a Lidia animal immersed in other populations or occupying different ecological niches, thus showing low ecological exchangeability (Crandall *et al.* 2000). In other words, the Lidia breed occupies an adaptive zone different from that of any other breed in its range and evolves separately from the breeds outside its range. Although there is a desirable pattern of aggressiveness present in the behaviour of the whole breed, different types of traditional popular events demand different types of behaviour. These variations in behavioural needs have been one of the main reasons why this racial group has fragmented into small lineages, traditionally called *encastes* (Boletín Oficial del Estado 2001), with different levels of gene flow among them and different behavioural traits favoured in each one. Some of the more isolated lineages, with a low number of animals contributing to the gene pool, can suffer from a loss of genetic variation and inbreeding depression, while others could be the result of a mixture of ancestral *encastes*.

Address for correspondence

J. Cañón, Department of Animal Production, Facultad de Veterinaria, Universidad Complutense, 28040 Madrid, Spain.
E-mail: jcanon@vet.ucm.es

Accepted for publication 7 April 2008

In summary, the distinct Mediterranean ecosystem inappropriate for most of the specialized bovine breeds, the presence of several lineages or *encastes* and the extreme aggressiveness are characteristics that strongly prevent the exchangeability (Crandall *et al.* 2000) of this breed with others.

Recently, the genetic position of this breed with respect to a collection of European cattle breeds has been addressed (see fig. 2 in Negrini *et al.* 2007) but a more precise knowledge of the structure of such a particular breed should be of interest to classify their lineages and identify the underlying amount and distribution of genetic variability, which can be used in the future to define conservation priorities in the fighting-bull breed.

Materials and methods

Sampling of populations

Detailed figures of the number of herds and animals sampled in each lineage or *encaste* are shown in Table 1. Some of the lineages are composed of a single herd, from which the samples for analysis were taken, e.g. Pablo Romero, Miura, Arauz de Robles or Cuadri. For each of the remaining lineages, a wide range of herds were selected and sampled.

For each lineage, individuals from the same generation were randomly sampled. Fresh blood collected in a buffer formulated to prevent DNA degradation (Dunner & Cañón

Lineage	Number of herds	Number of samples	H_E	H_O	Number of alleles	Allelic richness ¹	F_{IS}
Antonio Pérez	1	45	0.560	0.539	4.4	3.8	0.04
Arauz de Robles	1	52	0.536	0.530	4.8	3.9	0.01
Atanasio Fernández	6	97	0.552	0.511	5.0	3.8	0.07*
Baltasar Iban	2	52	0.577	0.534	4.7	4.0	0.07*
Braganza	1	25	0.587	0.575	4.4	–	0.02
Carlos Núñez	6	71	0.646	0.573	5.8	4.7	0.11*
Concha y Sierra	1	49	0.650	0.603	5.1	4.5	0.07*
Conde de la Corte	1	30	0.465	0.467	3.5	3.1	–0.003
Conde Santa Coloma	10	182	0.673	0.550	7.0	5.0	0.18*
Contreras	4	59	0.684	0.589	6.0	5.1	0.14*
Cuadri	1	50	0.464	0.430	4.3	3.4	0.07*
Félix Gómez	1	46	0.612	0.583	4.9	4.2	0.05*
Gamero Cívico	4	57	0.543	0.433	4.7	3.9	0.20*
Hidalgo Barquero	4	57	0.610	0.514	5.2	4.3	0.16*
José Marzal	1	50	0.600	0.590	5.3	4.4	0.02*
Juan Pedro Domecq	9	212	0.578	0.489	6.2	4.2	0.15*
Manuel Arranz	1	32	0.548	0.559	4.0	3.7	–0.02
Maria Montalvo	1	11	0.544	0.548	3.7	–	–0.008
Marqués de Albasarreda	3	46	0.528	0.480	4.2	3.4	0.09*
Marqués de Villamarta	3	60	0.610	0.517	5.3	4.4	0.15*
Miura	1	46	0.588	0.525	4.7	4.0	0.11*
Murube	4	51	0.564	0.463	4.8	4.1	0.18*
Pablo Romero	1	50	0.569	0.539	4.4	3.8	0.05*
Pedrajas	2	48	0.570	0.485	4.8	3.9	0.15*
Saltillo	4	54	0.652	0.507	5.5	4.6	0.22*
Torrestrella	1	50	0.571	0.570	4.8	4.1	0.001
Urcola	1	23	0.643	0.593	4.4	–	0.08*
Vega Villar	3	46	0.614	0.458	5.0	4.3	0.25*
Veragua	1	32	0.638	0.566	5.6	4.1	0.11*
Averaged across lineages	2.7	58.0	0.59	0.530	4.9	4.1	0.12*
All lineages	79	1683	0.715	0.524	9.8	5.7	0.27*
Avileña		50	0.764	0.688	8.0	6.5	0.10*
Morucha		50	0.764	0.725	8.0	6.6	0.05*
Retinta		50	0.780	0.707	8.6	6.8	0.09*
Averaged across breeds		50	0.769	0.707	8.2	6.6	0.08*
All local breeds		150	0.791	0.710	10.3	7.2	0.10*

*Values different from 0 ($P < 0.01$).

¹Based on a minimum sample size of 15 individuals.

Table 1 Description of 29 lineages of the Lidia breed.

2006) was taken from 1683 individuals, with approximately the same number of males and females.

Additionally, 50 animals from each of three local bovine breeds (Morucha, Retinta and Avileña) of similar geographical origin and with a production system similar to that of the Lidia breed were also included in the analysis as reference for some population genetic parameters.

Genetic loci

The 24 microsatellite loci studied, their chromosomal locations and the numbers of alleles in the population analysed are described in Table S1. Nineteen of these overlap with the panel used in the Resgen project CT98-118 'Towards a strategy for the conservation of the genetic diversity of European cattle', and the remaining five were considered of interest for this study because of the high variability found in previous analyses with this breed.

Statistical analysis

Genetic analyses were performed assuming the existence of traditional lineages or *encastes* to which the herds were assigned (Table 1). These assignments were made following the standards set by the UCTL breeders' association (Unión de Criadores de Toros de Lidia).

For each lineage or *encaste*, the observed heterozygosity, H_O , and unbiased expected heterozygosity, H_E , were estimated according to Nei (1978) and averaged over loci following Nei (1987) (Table 1). Allelic richness based on a minimum sample size of 15 individuals was calculated with the *ESTAT* v.2.9.3 program (<http://www2.unil.ch/popgen/softwares/estat.htm>) and averaged over the markers. The Braganza, María Montalvo and Urcola lineages were excluded from this analysis because of their low sample size.

Wright's F -statistics (F_{IT} , F_{IS} and F_{ST} , Wright 1951) were used to assess the distribution of genetic variability within and among lineages following Weir & Cockerham's (1984) approach. Departures from Hardy–Weinberg proportions in the whole sample and within each lineage were determined with Wright's F_{IT} and F_{IS} respectively. The null distributions of F_{IT} and F_{IS} were approximated by permuting alleles 1000 times within the whole set or within each lineage respectively. Random permutations of genotypes among samples were performed to test whether F_{ST} departed significantly from zero. These computations were performed using *ESTAT* and sequential Bonferroni (Rice 1989) was applied for multiple test correction.

Genetic relationships between lineages were inferred using two approaches. First, pairwise Reynolds' distances (Reynolds *et al.* 1983) were computed for all pairs of the 29 lineages and used to construct a neighbour-joining (NJ) tree with 1000 bootstrap resamplings to assess their consistency. *PHYLIP* software (<http://evolution.genetics.washington.edu/phylip.html>) was used to perform this analysis. These genetic

distances were also used to obtain a transformed matrix with ultrametric properties (Weitzman 1992, 1993), and a NJ tree for this new matrix was drawn using *MEGA* 2.0 software (Kumar *et al.* 2001). Because of the nature of this matrix, the tree is unique in the sense that any grouping algorithm (NJ, UPGMA, etc.) would return the same structure and this tree structure is the one that has the highest probability of explaining the presence of all the taxonomical units today (Weitzman 1992). Examples of the use of this kind of tree can be found in Thaon d'Arnoldi *et al.* (1998), Cañón *et al.* (2001), Barker *et al.* (2001) and Reist-Marti *et al.* (2003).

Second, a multivariate approach was considered. More specifically, a correspondence analysis (Benzécri 1973) was performed using *GENETIX* 4.04 (Belkhir *et al.* 2001) to plot the representation of lineages in a metric space.

Alternatively, the unsupervised method of Pritchard *et al.* (2000) implemented in the *STRUCTURE* program (<http://pritch.bsd.uchicago.edu/structure.html>) was used to assess the support that the molecular information offered to the *a priori* assumptions on the population structure. The whole set of animals was pooled into a single unit to analyse their underlying genetic structure. We examined clustering and whether or not these clusters corresponded to the assumed lineages. For the ancestry model, animals were allowed to have mixed ancestry, and burn-in and run-in lengths of 300 000 and 500 000 respectively were used. After determining the maximum likelihood value of the number of ancestral populations (31), the clustering success rate was assessed following the procedure described in fig. 1 of Rosenberg *et al.* (2001). Animals were then grouped into herds (some herds were pooled into one when they were similarly managed or when they were recently split from an initial herd after its owner's death) and pairwise distances were calculated following Cañón *et al.* (2006) to represent the relationships obtained by the *STRUCTURE* algorithm, which were then displayed in an NJ tree built with these distances.

Results

Within-lineage diversity and deviation from Hardy–Weinberg equilibrium

In total, 234 alleles were detected across the 24 loci and 1683 individuals sampled from 29 lineages. Average H_E across loci varied between 0.46 (Cuadri) and 0.68 (Contreras) (Table 1). Conde de la Corte, which had the second lowest heterozygosity, also had the lowest average number of alleles and allelic richness. In contrast, the Conde de Santa Coloma lineage possessed the largest number of alleles and Contreras, the largest allelic richness (Table 1). The reference local cattle breeds registered higher values for some of the diversity parameters, such as H_O and allelic richness, which were 33% and 60% higher respectively.

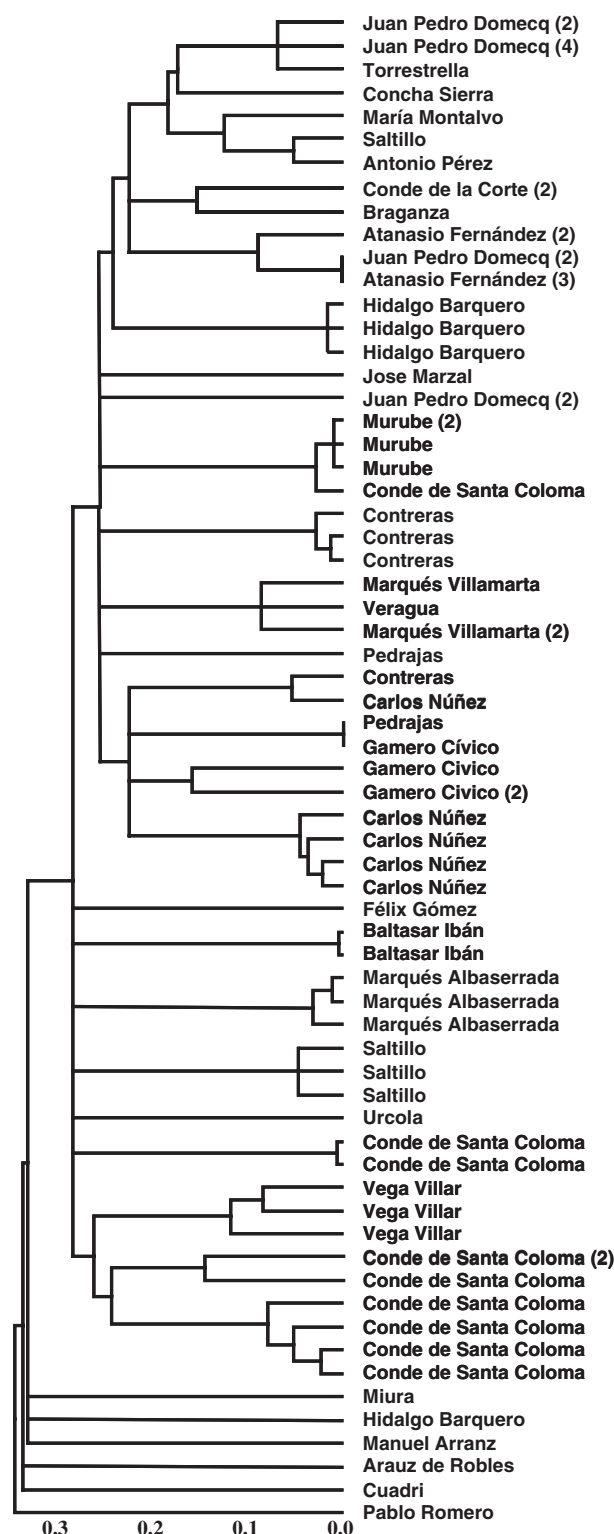


Figure 1 Neighbour-joining tree based on distances obtained from STRUCTURE output. The number of herds in each lineage is in parentheses.

Pearson correlation coefficients between heterozygosity and number of alleles and between number of alleles and allelic richness were 0.73 and 0.86 respectively.

The average F_{IT} value for all loci across populations (0.27) was highly significant [bootstrap 95% confidence interval (0.25, 0.30)] indicating a strong deficit of heterozygotes. This was caused both by significant departures from Hardy–Weinberg expectation within lineages [14 lineages exhibited significant across loci F_{IS} values, average $F_{IS} = 0.11$, bootstrap 95% confidence interval (0.09, 0.14)] and by differentiation between lineages [average $F_{ST} = 0.18$, bootstrap 95% confidence interval (0.17, 0.19)].

Differentiation between lineages

Genetic differentiation was high and significant with an average F_{ST} of 0.18. All pairwise F_{ST} values were highly significant, ranging from 0.05 (Torrestrella vs. Juan Pedro Domecq) to 0.35 (Marqués de Albaserrada vs. Conde de la Corte) indicating strong differentiation and reproductive isolation among the lineages. For each population, the average of Reynolds' distance with the rest of the lineages was calculated as an indicator of the divergence of the populations, and results arranged in descending order appear in Table S2. On average, the Cuadri lineage appeared as the most genetically isolated from the rest of the lineages, and Contreras, the least. The first two axes of the correspondence analysis, which contribute 12.9% and 9.4% of the total inertia respectively, are shown in Fig. S1. The local cattle breeds are separated from the Lidia lineages on axis 1.

Despite the high levels of genetic differentiation observed between the lineages, there was low statistical support for the neighbour-joining cluster analysis (data not shown) and only four branching events exhibited bootstrap support greater than 70%. This lack of a strong clustering pattern indicated that the exact relationship between lineages could not be reliably established with these 24 markers. The dendrogram in Fig. S2, obtained with the Reynolds' distances transformed into ultrametric distances, showed that the lineages grouped into two main clusters. The length of the branches corresponded to the level of inbreeding; therefore, the Cuadri, Arauz de Robles, Marqués de Albaserrada, Pablo Romero and Miura lineages appear to have undergone higher levels of inbreeding than the rest.

When performing cluster analysis with STRUCTURE, the different herds mostly grouped together as expected from *a priori* assignments of herds to lineages. The maximum likelihood number of clusters turned out to be 31, so the 29 lineages would actually come from 31 genetically differentiated ancestral groups. Pairwise distances were built following Cañón *et al.* (2006) to represent in a hierarchical tree the relationships of the different herds. The associated tree, showing how herds would group according to the STRUCTURE output, is depicted in Fig. 1. Table S2 shows the percentage of successful assignments (individuals assigned to their own lineage) obtained with the procedure of Rosenberg *et al.* (2001). The success rate was generally quite high.

Discussion

Variability and heterozygote deficits within lineages

Genetic diversity within lineages of the Lidia breed is, in general, lower than in European bovine breeds, but this does not extend to the results for the Lidia breed considered as a whole. The number of alleles per lineage ranged between 3.5 and 7, and the average number of alleles per locus observed in the Lidia breed (9.8) was very close to that observed in the three local cattle breeds included in the study (Table 1). Allelic richness (number of alleles standardized for variation in sample size) for the whole breed (5.7) was similar to the average obtained for a set of 69 breeds analysed by the European Cattle Genetic Diversity Consortium (2006) when the parameter is computed based on a minimum of 11 animals. However, it is low when compared to the average of the set of reference local breeds (5.7 vs. 7.2) (Table 1), although in this case the analysis was based on a minimum sample size of 15 animals. Observed heterozygosities within lineages ranged between 0.43 and 0.60, and the value for the whole breed was 0.52. These values are lower than the 0.69 average (range 0.52–0.73) of the 69 breeds considered by the European Cattle Genetic Diversity Consortium (2006).

Gene diversity of the Lidia breed is quite similar to that found in the three local breeds (Table 1), and similar expected heterozygosities were also found in the 69 European breeds analysed by the European Cattle Genetic Diversity Consortium (2006). Nonetheless, the average value within lineages was, as expected, lower than that of the European cattle breeds (0.59 vs. 0.71).

Divergence between gene diversity and H_0 shows up as a strong heterozygote deficit (high F_{IS} values) in 14 of the 29 lineages. The likely explanation for these heterozygote deficits is the sampling process within many of the lineages. Most of the lineages with high F_{IS} values included several herds (Table 1), which, in practice, can be genetically differentiated groups, and this stratification was thus present in the analysed sample. After reanalysing the data by taking the herd structure into consideration, the result was that only six of the 146 pairwise F_{ST} values were non-significant. The F_{ST} values within lineages varied between 0.05 (Baltasar Ibán) and 0.27 (Vega Villar) (Table S3), with an unweighted average of 0.15. Most of those F_{ST} values were greater than the average within-lineage F_{IS} value of 0.11, which means that genetic differences between herds were responsible for most, but not all, of the heterozygote deficit. The average F_{ST} value of 0.15 is greater than the average pairwise F_{ST} values obtained between bovine breeds in some recent publications (Cañón *et al.* 2001; European Cattle Genetic Diversity Consortium 2006), which were around 0.07. Although the use of assisted reproductive technologies is rare in the Lidia breed, breeding practices such as inbreeding could lead to small effective population sizes.

Actually, the effective number of founders, i.e. the number of equally contributing founders that would be expected to produce the same genetic diversity as that observed in the population under study (Lacy 1989), is below five for most lineages, and effective population size is, on average, around 30 (data not shown). Therefore, the positive F_{IS} values observed within the lineages and even in some herds are explained by the existence of genetic isolation among herds within the lineages and by some degree of inbreeding within herds.

Analysis of within-lineage genetic variability classified the lineages into three categories that almost perfectly matched their breeding practices. The first category comprised those lineages with one single herd, in which inbreeding practices are used to fix desired patterns of behavioural traits. Vera-gua, Urcola, Miura, Concha y Sierra and Pablo Romero are examples of single-herd lineages with relatively high inbreeding values ($F_{IS} > 0.07$). The second category included lineages with several herds and restricted migration among them. Most of the lineages in Table S3 belong to this category, and all of them showed high heterozygosity deficit levels ($F_{IS} > 0.15$) because of subdivision in genetically different herds. Finally, a third category grouped together lineages that contained either recent crossbreeding with animals from different lineages or recent pooling of different strains or families within the lineages. Antonio Pérez, Braganza, José Marzal, María Montalvo and Torrestrella are examples of the second category, while Arauz de Robles is an example of the third. All of them are characterized by a F_{IS} value that is not different from 0.

Differentiation between lineages and individuals

The level of differentiation between lineages was very large, as demonstrated by an average F_{ST} of 0.18. This value was much higher than values among European cattle breeds, which ranged from 0.07 (Cañón *et al.* 2001) to 0.10 (MacHugh *et al.* 1998). In fact, in the present study, all average F_{ST} values between each lineage and the rest of the lines were greater than 0.12 (Table S2).

Under a pure genetic drift model, assuming no admixture and no mutation, the Reynolds' distance between populations i and j and their effective sizes were used to estimate time divergence between the two populations using the formula $t \sim D_R / (1/4N_{ei} + 1/4N_{ej})$. On average, time since divergence in generations among lineages was around 13, equivalent to 95 years, for an average generation interval of 7.5 years. They ranged between 4.4 generations (~ 33 years), for the divergence between the Juan Pedro Domecq and Torrestrella lineages, and 17.5 generations (~ 137 years), for the divergence between the Conde de la Corte and Marqués de Albaserrada lineages. However, information recovered in the UCTL breeders' association from historical data indicated that at least for some pairs of lineages, longer divergence times were expected. For

instance, genetic isolation between the Conde de la Corte and Marqués de Albaserrada lineages has been documented since 1820.

Regarding the correspondence analysis, the most notable results are the very clear separation between the three local bovine breeds and all the lineages of the Lidia breed, and the recognition of most Lidia lineages using only the first two axes, a clear sign of the high degree of genetic differentiation.

The order of the lineages appearing in the tree based on Weitzman's diversity function (Fig. S1) depended on the random drift that caused the observed diversity. On the other hand, despite the lack of statistical support for the tree topology, the accuracy of the assignment of individuals to their own lineage was high (95% on average) and in 13 of the 29 lineages, 100% of the individuals were correctly assigned (Table S2).

Herd clustering into the 31 groups obtained from the STRUCTURE analysis fitted the *a priori* assumptions. There were some mismatches, mostly arising out of three situations: (i) some lineages, initially considered as different, were now clustered together, like Antonio Pérez and María Montalvo, Braganza and Conde de la Corte, Torrestrella and some herds from Juan Pedro Domecq or Veragua and Marqués de Villamarta; (ii) herds belonging to the same lineage now appeared clearly separated, such as those from Conde de Santa Coloma, divided now into three different groups, or those from Juan Pedro Domecq and Hidalgo Barquero, split into two groups each; and (iii) herds from different lineages were joined together into a new lineage, as happens with one herd from Pedrajas and one from Gamero Cívico, or one from Carlos Núñez and one from Contreras. Lineages have often been generated from a low number of animals that shared a particular phenotype (usually behavioural traits). Breeders have also frequently tried to maintain the desired characteristics by mating related animals (i.e. inbreeding). Lineages are traditionally kept within families, which hinder gene flow and rapidly establish significant genetic differentiation between lineages. This fragmentation of the fighting-bull population, and consequently the low population sizes of its lineages, seriously compromises the short-term survival of lineages and the loss of their specific genes. On the other hand, population subdivision can be a powerful way of maintaining genetic variability, because different alleles at a locus can be fixed in different lineages and be permanently kept as a source of variation, as long as the lineage persists. In fact, assuming no mutation, migration or selection, the amount of genetic variance preserved among n lineages approaches $(1 - 1/n)V_g(0)$ after few times N_e/n generations, with $V_g(0)$ being the genetic variance at coalescence and N_e the effective population size of the breed (Lande 1995).

In situations of subdivision and lack of genetic connections between lineages, the use of molecular information to estimate between-population kinship has shown an acceptable efficiency in terms of predicting low error (Eding & Meuwissen 2001). On the other hand, there is very detailed

pedigree information within each lineage. Thus, methods that combine between- and within-lineage variation could be used to establish conservation of genetic diversity in the fighting-bull breed (Eding *et al.* 2002; García *et al.* 2005; Ollivier & Foulley 2005), where the between-lineage variability is estimated using the molecular information (Weitzman 1992, 1993; Eding & Meuwissen 2001), and within-lineage variation is based on available pedigree data.

Acknowledgements

The project has been funded by CICYT and European Regional Development Fund no. 2FD97-1191, and also by the INIA and European Regional Development Fund no. RZ01-002-C2-1, RZ01-002-C2-2 and RZ2004-00009. Thanks are also due to the veterinarians of the UCTL and all the breeders participating who provided the blood samples.

References

- Barker J.S.F., Tan S.G., Moore S.S., Mukherjee T.K. & Matheson J-L. (2001) Genetic variation within and relationships among populations of Asian goats (*Capra hircus*). *Journal of Animal Breeding and Genetics* **118**, 213–33.
- Belkhir K., Borsa P., Chikhi L., Raufaste N. & Bonhomme F. (2001) *Genetix, logiciel sous Windows TM pour la génétique des populations*. Laboratoire Génome, Populations, Interactions, CNRS UPR 9060. Université de Montpellier II, Montpellier (<http://www.univ-montp2.fr/~genetix/genetix/genetix.htm>).
- Benzécri J.P. (1973) *Analyse des données. Tome 2: Analyse des correspondances*. Dunod, Paris.
- Blott S.C., Williams J.L. & Haley C.S. (1998) Genetic variation within the Hereford breed of cattle. *Animal Genetics* **29**, 202–11.
- Boletín Oficial del Estado (2001) Real Decreto 60/2001, de 26 de enero, sobre prototipo racial de la raza bovina de lidia. *Boletín Oficial del Estado* **38**, 5255–61.
- Buchanan F.C., Adams L.J., Littlejohn L.P., Maddox J.F. & Crawford A.M. (1994) Determination of evolutionary relationships among sheep breeds using microsatellites. *Genomics* **22**, 397–403.
- Cañón J., Alexandrino P., Bessa I. *et al.* (2001) Genetic diversity measures of local European beef cattle breeds for conservation purposes. *Genetics Selection Evolution* **33**, 311–32.
- Cañón J., García D., García-Atance M.A., Obexer-Ruff G., Lenstra J.A., Ajmone-Marsan P., Dunner S. & the ECONOGENE Consortium (2006) Geographical partitioning of goat diversity in Europe and the Middle East. *Animal Genetics* **37**, 327–34.
- Crandall K.A., Bininda-Emonds O.R.P., Mace G.M. & Wayne R.K. (2000) Considering evolutionary processes in conservation biology. *Trends in Ecology and Evolution* **15**, 290–5.
- Dunner S. & Cañón J. (2006) Patent WO 2006/040376.
- Eding J.H. & Meuwissen T.H.E. (2001) Marker based estimates of between and within population kinships for the conservation of genetic diversity. *Journal of Animal Breeding and Genetics* **118**, 141–59.
- Eding H., Crooijmans R.P.M.A., Groenen M.A.M. & Meuwissen T.H.E. (2002) Assessing the contribution of breeds to genetic diversity in conservation schemes. *Genetics Selection Evolution* **34**, 613–33.

- European Cattle Genetic Diversity Consortium (2006) Marker-assisted conservation of European cattle breeds: an evaluation. *Animal Genetics* **37**, 475–81.
- García D., Corral N. & Cañón J. (2005) Combining inter- and intrapopulation information with the Weitzman approach to diversity conservation. *Journal of Heredity* **96**, 704–12.
- Kumar S., Tamura K., Jakobsen I.B. & Nei M. (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**, 1244–5.
- Lacy R. (1989) Analysis of founder representation in pedigrees: founder equivalents and founder genome equivalence. *Zoo Biology* **8**, 111–24.
- Lande R. (1995) Breeding plans for small populations based on the dynamics of quantitative variance. In: *Population Management for Survival and Recovery: Analytical Methods and Strategies in Small Population Conservation* (Ed. by J. D. Ballou, M. Gilpin & T.J. Foose), pp. 318–40. Columbia University Press, New York.
- MacHugh D.E., Loftus R.T., Cunningham P. & Bradley D.G. (1998) Genetic structure of seven European cattle breeds assessed using 20 microsatellite markers. *Animal Genetics* **29**, 333–40.
- Negrini R., Nijman I.J., Milanesi E. *et al.* (2007) Differentiation of European cattle by AFLP fingerprinting. *Animal Genetics* **38**, 60–6.
- Nei M. (1978) Estimation of heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**, 583–90.
- Nei M. (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- Ollivier L. & Foulley J.L. (2005) Aggregate diversity: new approach combining within- and between-breed genetic diversity. *Livestock Production Science* **95**, 247–54.
- Pritchard J.K., Stephens M. & Donnelly P. (2000) Inference of population structure from multilocus genotype data. *Genetics* **155**, 945–59.
- Reist-Marti S.B., Simianer H., Gibson J., Hanotte O. & Rege J.E.O. (2003) Weitzman's approach and conservation of breed diversity: an application to African cattle breeds. *Conservation Biology* **17**, 1299–311.
- Reynolds J., Weir B.S. & Cockerham C.C. (1983) Estimation of the coancestry coefficient: Basis for a short-term genetic distance. *Genetics* **105**, 767–79.
- Rice W.R. (1989) Analyzing tables of statistical tests. *Evolution* **43**, 223–5.
- Rosenberg N.A., Burke T., Elo K. *et al.* (2001) Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. *Genetics* **159**, 699–713.
- Silva B., Gonzalo A. & Cañón J. (2002) Genetic parameters of behavioural traits in the bovine (*Bos taurus*). *Proceedings of the 7th World Congress on Genetics Applied to Livestock XXXII*, 83–6.
- Silva B., Gonzalo A. & Cañón J. (2006) Genetic parameters of aggressiveness, ferocity and mobility in the fighting bull breed. *Animal Research* **55**, 65–70.
- Thaon d'Arnoldi C., Foulley J.L. & Ollivier L. (1998) An overview of the Weitzman approach to diversity. *Genetics Selection Evolution* **30**, 149–61.
- Weir B.S. & Cockerham C.C. (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**, 1358–70.
- Weitzman M. (1992) On diversity. *Quarterly Journal of Economy* **107**, 363–405.
- Weitzman M. (1993) What to preserve? An application of diversity theory to crane conservation. *Quarterly Journal of Economy* **108**, 157–83.
- Wright S. (1951) The genetical structure of populations. *Annals of Eugenics* **15**, 323–54.

Supplementary material

The following supplementary material is available for this article online from <http://www.blackwell-synergy.com/doi/full/10.1111/j.1365-2052.2008.01738.x>

Figure S1 Projection of 29 lineages of the Lidia breed plus three local breeds (Morucha, Retinta and Avileña) on axes 1 and 2 using the coordinates obtained by a correspondence analysis.

Figure S2 Neighbour-joining tree based on Reynolds' genetic distances transformed to distances with ultrametric properties.

Table S1 Microsatellite names, numbers of alleles and chromosomal locations.

Table S2 Average Reynolds' genetic distances from each lineage to the rest classified in descent order and clustering success rate.

Table S3 Wright's *F*-statistics within each lineage.

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors.

Ancestral matrilineages and mitochondrial DNA diversity of the Lidia cattle breed

O. Cortés*, I. Tupac-Yupanqui*, S. Dunner*, M. A. García-Atance*, D. García*, J. Fernández[†] and J. Cañón*

*Department of Animal Production, Facultad de Veterinaria, Universidad Complutense, 28040 Madrid, Spain. [†]Unión de Criadores de Toros de Lidia, Departamento Técnico, Paseo de Eduardo Dato 7, 28010 Madrid, Spain

Summary

To clarify the genetic ancestry and the mitochondrial DNA (mtDNA) diversity of the Lidia cattle breed, a 521-bp D-loop fragment was sequenced in 527 animals belonging to 70 herds distributed across 29 lineages. The mtDNA diversity recorded was similar to that seen for Middle Eastern breeds and greater than that recorded for the majority of European breeds. Haplotype T3 was the most common (81%), followed by the African T1 haplotype (17%); very low frequencies were recorded for haplotypes T and T2. The results agree with there being two major ancestral lines for the Lidia breed, European and African, similar to that seen for other Mediterranean breeds. A wide range of variation in haplotype frequencies was seen between the examined lineages. Haplotype T3 was present in all those analysed; in five it was the only one present, and in only one lineage (Miura) was its frequency lower than that of T1. T1*, a haplotype reported in Criollo breeds and to date in only a single European breed (the Retinta breed from Spain), was found in a single animal belonging to the Concha y Sierra lineage. Network analysis of the Lidia breed revealed the presence of two major haplotypes: T3 and T1. The Lidia breed appears to be more closely related to prehistoric Iberian and Italian than to British aurochs.

Keywords aurochs, bovine, fighting bull, maternal lineage, mitochondrial DNA.

Introduction

The analysis of mitochondrial DNA (mtDNA) sequence diversity has provided useful information on the origin and diversification of current cattle populations (Loftus *et al.* 1994; Bradley *et al.* 1996; Troy *et al.* 2001). Wild aurochs were domesticated some 12 000 years ago across Europe, Asia and Africa, and the mitochondrial signals of these animals can be seen in modern cattle breeds (Loftus *et al.* 1994; Troy *et al.* 2001; Edwards *et al.* 2004). Five main mitochondrial haplotypes have been described, T, T1, T2, T3 and T4, which show different geographical distributions (Cymbron *et al.* 1999; Troy *et al.* 2001). Haplotype T3 is clearly predominant in central and northern European breeds, while T1 is mainly seen in African breeds (Beja-Pereira *et al.* 2006). Mediterranean breeds have both the T3 and T1 (at a lower frequency) haplotypes because of African

influence. Haplotypes T and T2 appear in the breeds of Europe and Africa but at very low frequencies (Bradley *et al.* 1996; Cymbron *et al.* 1999; Troy *et al.* 2001), and T4 is found in Asian breeds (Mannen *et al.* 2004). A new haplotype, T1* (defined as T1 with mutations at positions 16053, 16122, 16139 and 16196) has been reported in Criollo breeds and (to date) in a single European breed (the Retinta breed from Spain) (Miretti *et al.* 2002, 2004) (Table S1).

The relationships between wild aurochs and domesticated cattle breeds, however, remain unclear (Vilà *et al.* 2005). The first auroch mtDNA sequences, collected in Great Britain, typed far from those of modern cattle breeds, suggesting little auroch introgression (Troy *et al.* 2001). Later, however, more ancient auroch sequences from Italy and from the Bronze Age of the Iberian Peninsula revealed haplotype distributions similar to those of modern European cattle breeds (Anderung *et al.* 2005; Beja-Pereira *et al.* 2006). Only one Iberian sample appeared more closely related to the British auroch sequences (Anderung *et al.* 2005). Thus, the introgression of auroch mtDNA into modern cattle breeds has taken place, but it is not clear to

Address for correspondence

J. Cañón, Department of Animal Production, Facultad de Veterinaria, Universidad Complutense, 28040 Madrid, Spain.
E-mail: jcanon@vet.ucm.es

Accepted for publication 29 June 2008

what degree or whether this varied depending upon geographical location.

Three waves of migration from the original area of cattle domestication arrived at the Iberian Peninsula, influencing the then-current cattle breeds. Two of these migrations followed the route from the Middle East via the European mainland or Mediterranean coast, while the third took the African route; this was possible because of the proximity between Iberia and Africa (Cymbron *et al.* 1999; Anderung *et al.* 2005). African populations have entered the Iberian Peninsula since the Bronze Age (Bogucki 1996; Gkliasta *et al.* 2003). Thus, Spanish cattle breeds may show the northern or southern European haplotype pattern depending on their origin.

The Lidia breed, otherwise known as the fighting bull, has a number of peculiarities that make it one of the most successful of domestic breeds in Spain. It has been spread throughout this country, Portugal and southern France, as well as throughout numerous Central and South American countries (<http://www.toroslidia.com/>). The first known reference to a Lidia herd is found in the 14th century literature of the kingdom of Navarra: "Joan de Gris, breeder of Tudela, sent in 1388 two bulls to Pamplona by order of King Charles III 'The Noble' for the fair, in honor of the Duke of Borbon" (Larrea *et al.* 2005). Between the eighth and 15th centuries, a constantly changing no-man's land formed the frontier between the Muslim and Christian regions of the Iberian Peninsula, favouring the existence of a semi-feral bovine population. The most feral, aggressive animals were commonly selected for fiestas.

During the 18th century, a certain amount of breeding specialization took place, and animals from the traditional herds, which mainly produced meat, were selected to produce animals for these social events. Different types of traditional spectacle demanded different types of behaviour on the part of the bull, which largely explains why this racial group has become fragmented into small lineages, traditionally called *encastes* (Boletín Oficial del Estado 2001), with different levels of gene flow among them. As a consequence, the Lidia breed shows great phenotypic (morphological and behavioural) heterogeneity among its lineages (Boletín Oficial del Estado 2001). The present-day representatives of these lineages are now mainly found in the west and southwest of the Iberian Peninsula, and the lineages do not follow any particular distribution pattern across the geographical distribution areas.

The Lidia breed has likely been subjected to two major historical influences, one exerted by cattle that accompanied the Celtic invasions, and one exerted by African populations, a consequence of geographic proximity (<http://www.toroslidia.com/>). However, the few DNA studies performed have been insufficient to corroborate this idea, and none have examined mtDNA.

Despite the peculiarities described above and the great genetic diversity manifested in its nuclear genetic material

(Cañón *et al.* 2008), little is known about the genetic influences that helped form the Lidia breed. The aim of the present work was to investigate these influences by examining the mtDNA of its lineages.

Materials and methods

Five hundred and twenty-nine blood samples were collected from the animals bred in Spain and maintained in a DNA preservation buffer (Dunner & Cañón 2006) until use. Table S2 shows the number of herds and unrelated animals sampled from each lineage. DNA was extracted by standard methods (Sambrook *et al.* 1989). A 521-bp fragment of the mtDNA D-loop was sequenced, encompassing positions 16019–16201 (Anderson *et al.* 1982). Two overlapping D-loop fragments were amplified using PCR. Primer sequences were designed using the software PRIMER version 0.5 (GCG software package). The primers used for the first fragment were DLOOPF 5'-TATGCCCATGCATATAAGC-3' and DLOOPR 5'-AAGAAAGAACCAGATGCCTG-3'; the primers used for the second fragment were MITFOR 5'-ATCCCTCTTCTCGCTCCG-3' and MITREV 5'-TTATGTC-CTGTGACCATTGACTG-3'. The amplified fragments were purified using the Concert Rapid PCR Purification System (Life Technologies) according to the manufacturer's instructions. Sequencing was performed using primers DLOOPR and MITREV in an ABI PRISM 310 Genetic Analyzer (Applied Biosystems). Sequence alignment was performed using free software available at <http://bioinfo.hku.hk/EMBOSS/> (the European Molecular Biology Open Software Suite). The mean number of pairwise differences (MNP), nucleotide diversity, F_{ST} distances, and Fisher's exact test were obtained using MEGA software v2.1 (Kumar *et al.* 2001) and the ARLEQUIN package v3.11 (Excoffier *et al.* 2005, <http://cmpg.unibe.ch/software/arlequin3>). Reduced median networks (Bandelt *et al.* 1999) were generated with the resultant sequences and previously published *Bos primigenius* sequences (Troy *et al.* 2001; Anderung *et al.* 2005; Beja-Pereira *et al.* 2006) using the NETWORK 3.0 program (<http://www.fluxus-technology.com/>). Comparisons were restricted to the region of overlap between the Lidia and *Bos primigenius* D-loop mitochondrial sequences, between positions 16042–16280 (Anderson *et al.* 1982). Partitioning of the total genetic variance into components because of inter-lineages, inter-breeders and inter-individual differences was carried out using the analysis of molecular variance (AMOVA) program implemented in the ARLEQUIN package.

Results

The 529 sample sequences revealed 121 haplotypes, defined by 73 polymorphisms (65 transitions and eight transversions). A total of 53 haplotypes were unique, while the most common haplotypes were shared by 65 animals (frequency

of 12%). Six new mutations of the Old World haplotypes were detected (Table S1) at low frequencies (ranging from 0.2% to 1.3%) (EU373501–EU373506). The haplotype distribution for the Lidia breed as a whole showed a typical southern European pattern, with T3 the predominant haplotype (87%), T1 less common (17%), and T and T2 at very small frequencies (1.2% and 0.6% respectively) (Table 1). The frequency of the T1 haplotype ranges from 5% to 30% in different Mediterranean breeds (Beja-Pereira *et al.* 2006). In the Concha y Sierra lineage, one animal showed the Criollo haplotype (T1*); this has previously been described in just one European breed (Retinta from Spain) (Miretti *et al.* 2002, 2004).

All the lineages investigated showed the T3 haplotype; in five it was the only one present, and only in the Miura lineage did its frequency fall below 50%. Haplotype T1 was found at different frequencies in 24 of the 29 lineages examined, ranging from 3% in Atanasio Fernández to 57%

in Miura. Haplotype T2 was present in only two lineages and haplotype T in five, both at very low frequencies (Table 1). Fisher's exact test ($P = 0.006$) confirms the lack of randomness in the differences between lineages of the haplotype frequencies.

The mean MNPD value was 3.7, ranging between 2.3 in Antonio Pérez (which only showed haplotype T3) and 5.5 in Concha y Sierra (the only lineage to show four of the six haplotypes known to date; Table 1). The nucleotide diversity for the entire Lidia breed was 0.74%, varying from 0.4% in Antonio Pérez to 1% in Concha y Sierra (Table 1).

The pairwise F_{ST} values ranged from 0% to 30% (Pedrajas and Miura respectively), 74% of which were significant ($P < 0.05$). The average F_{ST} distance between each lineage and the others showed marked differences, ranging from 2.7% for Braganza to 17% for Pedrajas. Six lineages (Table 2) showed F_{ST} values of $>10\%$, indicating their reproductive isolation and strong differentiation. Six other lineages showed F_{ST} values of $<5\%$, indicating just the opposite. The remaining lineages showed intermediate situations, with F_{ST} values ranging from 5% to 10%. These high F_{ST} values are reflected in the high proportion of the

Table 1 Mean number of pairwise differences (MNPD), nucleotide diversity and frequencies of the main haplotypes for each lineage.

Lineage	MNPD ¹	Nucleotide diversity ²	Haplotypes (%)				
			T3	T1	T	T2	T1*
Pedrajas	2.6	0.005	100				
Antonio Pérez	2.3	0.004	100				
Urcola	3.4	0.007	100				
Pablo Romero	3.8	0.007	100				
María Montalvo	2.6	0.005	100				
Atanasio Fernández	2.7	0.005	95	2.5	2.5		
Baltasar Ibán	2.5	0.005	93	7			
Vega Villar	4.1	0.008	92	8			
Félix Gómez	4.0	0.008	87	13			
Torrestrella	3.3	0.006	87	13			
Veragua	3.9	0.007	86	14			
Juan Pedro Domecq	3.7	0.007	85	15			
Gamero Cívico	3.2	0.006	83	17			
Conde Santa Coloma	3.8	0.007	81	15	4		
José Marzal	3.9	0.008	80	13	7		
Arauz de Robles	4.1	0.008	80	20			
Contreras	3.4	0.007	80	7		13	
Marqués de Albaserrada	4.6	0.008	80	20			
Conde de la Corte	4.6	0.009	79	21			
Marqués de Villamarta	3.9	0.008	77	23			
Saltillo	3.6	0.007	75	25			
Carlos Núñez	3.4	0.007	75	20	5		
Hidalgo Barquero	3.6	0.007	71	29			
Braganza	3.9	0.008	71	29			
Murube	3.7	0.007	68	32			
Cuadri	3.6	0.007	67	33			
Manuel Arranz	5.0	0.010	64	29	7		
Concha y Sierra	5.5	0.011	56	31		5.5	5.5
Miura	4.0	0.008	43	57			
Average values	3.7	0.007	81	17	1.2	0.6	0.2

¹Standard errors ranged from 1.3 to 2.8.

²Standard errors ranged from 0.003 to 0.006.

Table 2 Average F_{ST} value for each lineage with respect to the others (descending order).

Lineage	F_{ST} distances (%)
Pedrajas	16.7
Miura	14.5
Concha y Sierra	12.7
Baltasar Ibán	12.5
María Montalvo	11.7
Urcola	9.6
Antonio Pérez	9.1
Vega Villar	8.5
Félix Gómez	8.1
Atanasio Fernández	8.0
Manuel Arranz	7.6
Contreras	7.6
Veragua	7.0
Pablo Romero	7.0
Cuadri	6.9
Arauz de Robles	6.8
Torrestrella	6.8
Hidalgo Barquero	6.3
Gamero Cívico	6.0
Murube	5.9
Conde de la Corte	5.8
Saltillo	5.5
Juan Pedro Domecq	4.5
Carlos Núñez	4.4
Marqués de Villamarta	4.3
Conde Santa Coloma	4.2
Marqués de Albaserrada	3.9
José Marzal	3.5
Braganza	2.7

Table 3 Partitioning of the genetic variability among the different sources of variation.

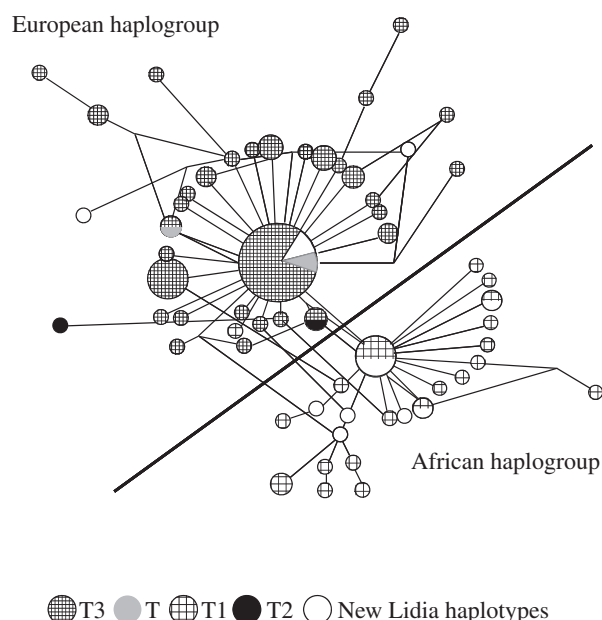
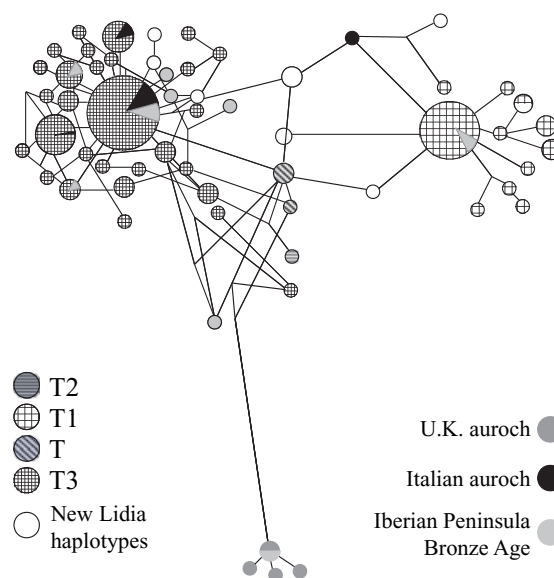
Source of variation	Percentage of variation	Fixation indices
Among lineages	1.2	$F_{SC} = 0.10$
Among breeders within lineage	10.1	$F_{ST} = 0.11$
Within breeders	88.8	$F_{CT} = 0.01$

total variance explained by the genetic differences among lineages obtained by the AMOVA analysis (Table 3).

Figure 1 shows the phylogeny network for the 121 haplotypes (Bandelt *et al.* 1999). Two main clusters can be recognized corresponding to the European haplogroup (T3 plus T and T2), and the African haplogroup (T1). Both networks show a star-like skeleton; the centre of the star corresponds to the haplotype with the highest frequency, while a number of peripheral haplotypes decoalesce. A similar network figure has been observed for Mediterranean breeds (Anderung *et al.* 2005). As described in previous studies, the T1* haplotype appeared in the African cluster but clearly separated from the centre.

The status of the six new haplotypes (Table S1) is dependent on nucleotide 16113. Those with the same nucleotide as haplotype T1 fall into the African cluster, while the rest fall into the European cluster.

When the auroch sequences from Great Britain, Italy and the Bronze Age Iberian samples were added to the network analysis, the topology showed four clusters (Fig. 2). The European and the African clusters still showed a star-like skeleton network. The British

**Figure 1** Reduced median-joining network (Bandelt *et al.* 1999) constructed from 121 Lidia breed haplotypes. Circle size is proportional to sequence frequency.**Figure 2** Reduced median-joining network (Bandelt *et al.* 1999) constructed from 121 Lidia breed haplotypes and auroch (*Bos primigenius*) samples from the UK (12300–11000 BCE), Italy (17100–11420 BCE) and the Iberian Peninsula (2670 BCE–1500 CE). Circle size is proportional to sequence frequency.

sequences and one Iberian sample formed another cluster, clearly isolated from the others. Finally, the fourth cluster was located between the European and African clusters, representing the T haplotype and the six new haplotypes found for the Lidia breed. This caused the loss of the star-like skeleton network, giving rise to a dispersed cluster in which no single sequence predominated. The Italian and all but one of Iberian auroch samples were represented in the T3 and T1 clusters.

Discussion

It is remarkable that the average MNPD for the Lidia breed is markedly higher than those of other European (1.8) or African breeds (2.1), but close to those of Middle Eastern (4.0) and Anatolian breeds (3.7) (Bradley *et al.* 1996; Troy *et al.* 2001; Carvajal-Carmona *et al.* 2003). The reduced diversity of European domestic cattle compared with those of the Middle East is consistent with the origin of these animals in the latter region. Nevertheless, it is unusual that a European breed should show an index of diversity similar to that of breeds from Anatolia and the Middle East. Analogous evidence can be obtained from the Lidia breed's nucleotide diversity; the value for the whole breed (0.74%) is higher than those of other European breeds, which range from 0.11% to 0.57% (Loftus *et al.* 1994). This may have two explanations: (i) the geographical location of the Iberian Peninsula and the three influences exerted during the breed-making process (the African, Mediterranean and mainland European influences) could have left the Lidia

breed in possession of all four major Old World haplotypes and (ii) the division of the breed into lineages with different levels of reproductive isolation may have favoured genetic differentiation, thus maintaining the breed's ancestral genetic diversity.

A high level of mitochondrial diversity was also seen within the different lineages studied. The lowest MNP value, of Antonio Perez (2.5), is higher than that of most European breeds. Only five lineages had an MNP value of <3. In three of these, only the T3 haplotype was present, and in the other two, it had a frequency of >93%. Eight lineages had an MNP value of >4, and with the exception of Vega Villar, the remainder showed a relatively high frequency for the T1 haplotype (>10%). The Pearson coefficient correlation between nucleotide diversity and the MNP value was 0.99%.

Diversity was particularly high in the Concha y Sierra lineage (MNP = 5.2; nucleotide diversity = 1.1%), in which three of the four Old World haplotypes were present. The identification of a T1* haplotype in a Concha y Sierra animal is remarkable; in Europe it was previously known only in the Spanish Retinta breed (Miretti *et al.* 2002, 2004; Liron *et al.* 2006). This finding is consistent with the presence of this lineage in southern North America and in Central America (a fact confirmed by historical documents). The presence of the T1* haplotype in the Concha y Sierra lineage may be explained by crosses in this area of the Concha y Sierra lineage with Criollo breeds previously introgressed with African zebu, the original carrier of the T1* haplotype (Liron *et al.* 2006).

The investigation of the nuclear DNA of the studied lineages has shown them to have a high degree of genetic differentiation (Cañón *et al.* 2008). The present work with mtDNA confirms this; indeed, and as opposed to what usually happens in domestic bovine breeds (Robertson 1953), the proportion of genetic variability caused by lineages (~11%) and even the genetic differentiation among breeders within lineages (~10%) were high (Table 3). It should be remembered, however, that the present results correspond to one single breed divided into several lineages – comparisons with the results obtained in studies involving several breeds should therefore be undertaken with caution. Among the lineages with the highest F_{ST} values ($F_{ST} > 9\%$), two groups can be identified. The first includes lineages with only one Old World haplotype, or with several but with one at a high frequency plus a lower MNP value (e.g. Pedrajas, Baltasar Iban, Antonio Pérez, María Montalvo and Urcola). In fact, the common practice of mating related animals with the objective of fixing desirable behavioural traits increases the effect of genetic drift, with genetic diversity becoming lost but the isolation between lineages increasing. The lineages of the second group (Miura and Concha y Sierra) have distinctive features that explain their genetic differentiation from the rest. Concha y Sierra show three distinct Old World hap-

lotypes plus T1*, while Miura is the only lineage with a T1 frequency higher than that of T3.

Two major influences on the Lidia breed have been described in historical documents, one exerted by the cattle that came with Celtic invaders, and the other by African populations, a consequence of their geographic proximity (<http://www.toroslidia.com>). Certainly, the present network skeleton analysis reveals the presence of two main matrilineages corresponding to European (T3) and African (T1) influences; this reinforces the idea that the Lidia breed is the result of a mixing with animals that arrived in two independent migration waves, one from Europe and one from Africa. The pattern shared by many European breeds, in which haplotype T3 is clearly more abundant than T or T2, confirms the influence exerted on the Lidia breed by domestic cattle originally from the Middle East that arrived in the Iberian Peninsula via the European mainland or Mediterranean route.

Some authors indicate that contemporary Iberian breeds are more closely related to African cattle than to European cattle because of gene flow from Africa in the 1960s and 1970s (Anderung *et al.* 2005). However, the Lidia herdbook indicates this as unlikely. Thus, the influence on the Lidia breed exerted by the African cattle might be traced back to the Muslim/North African Berber expansion (Cymbron *et al.* 1999) – or even earlier, as a recent study of a T1 haplotype in Iberian Bronze Age samples (Anderung *et al.* 2005) suggests.

European aurochs were probably strongly geographically structured, as the analysis of British and Italian aurochs indicates. However, a certain level of spreading may also have been possible given the finding in Spain of a Bronze Age sample with similarities to British aurochs. As expected, a network analysis including samples from prehistoric wild aurochs and prehistoric, domestic, Iberian samples showed more genetic affinity between the Lidia breed and Italian or prehistoric Iberian aurochs than with British aurochs.

One of the six new haplotypes for the Lidia breed is the same as that seen in an Italian auroch sample (haplotype 5 in Table S1). Nevertheless, its frequency in the Lidia breed is clearly lower than that in other Iberian breeds and closer to the frequencies of Balkan or mainland European breeds. It should be remembered that the few auroch samples analysed may not reflect the real distribution of auroch haplotypes. In addition, the similar skeleton networks shown by the Iberian Bronze Age samples and the Lidia breed indicate that no great changes have occurred in the mtDNA of the Lidia breed since the Bronze Age.

In conclusion, the high mtDNA diversity observed in the Lidia breed – closer to that seen in Middle Eastern than European populations – is evidence of a certain degree of primitivism. In fact, the Lidia breed was found to have all the Old World haplotypes plus T1*. The multiple influences on this breed throughout history and its division into

isolated reproductive lineages might explain the high level of mtDNA diversity seen.

Acknowledgements

We thank A. Burton for providing language advice. The project has been funded by CICYT and European Regional Development Fund no. 2FD97-1191, and also by the INIA and European Regional Development Fund nos RZ01-002-C2-1, RZ01-002-C2-2 and RZ2004-00009. Thanks are also due to the veterinarians of the UCTL and all the breeders who provided the blood samples.

References

- Anderson S., de Bruijn M.H., Coulson A.R., Eperon I.C., Sanger F. & Young I.G. (1982) Complete sequence of bovine mitochondrial DNA. Conserved features of the mammalian mitochondrial genome. *Journal of Molecular Biology* **156**, 683–71.
- Anderung C., Bouwman A., Persson P., Carretero J.M., Ortega A.I., Elburg R., Smith C., Arsuaga J.L., Ellegren H. & Götherström A. (2005) Prehistoric contacts over the straits of Gibraltar indicated by genetic analysis of Iberian Bronze Age cattle. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 8431–5.
- Bandelt H.J., Forster P. & Röhl A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**, 37–48.
- Beja-Pereira A., Caramelli D., Lalueza-Fox C. *et al.* (2006) The origin of European cattle: evidence from modern and ancient DNA. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 8113–8.
- Bogucki P. (1996) The spread of early farming in Europe. *American Science* **84**, 242–53.
- Boletín Oficial del Estado (2001) Real Decreto 60/2001, de 26 de enero, sobre prototipo racial de la raza bovina de lidia. *Boletín Oficial del Estado* **38**, 5255–61.
- Bradley D.G., MacHugh D.E., Cunningham P. & Loftus R.T. (1996) Mitochondrial diversity and the origins of African and European cattle. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 5131–5.
- Cañón J., Tupac-Yupanqui I., García-Atance M.A., Cortés O., García D., Fernández J. & Dunner S. (2008) Genetic variation within the Lidia bovine breed. *Animal Genetics* **39**, 439–45.
- Carvajal-Carmona L.G., Bermudez N., Olivera-Angel M., Estrada L., Ossa J., Bedoya G. & Ruiz-Linares A. (2003) Abundant mtDNA diversity and ancestral admixture in Colombian criollo cattle (*Bos taurus*). *Genetics* **165**, 1457–63.
- Cymbron T., Loftus R.T., Malhiero M.I. & Bradley D.G. (1999) Mitochondrial sequence variation suggests an African influence in Portuguese cattle. *Proceedings of the Royal Society of London. Series B* **266**, 597–603.
- Dunner S. & Cañón J. (2006) Solution for the indefinite maintenance of nucleic acids in the cell of origin thereof. Patent WO 2006/040376.
- Edwards J.E., MacHugh D.E., Dobney K.M. *et al.* (2004) Ancient DNA analysis of 101 cattle remains: limits and prospects. *Journal of Archaeological Science* **31**, 695–710.
- Excoffier L., Laval G. & Schneider S. (2005) ARLEQUIN version 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**, 47–50.
- Gkliasta M., Russel T., Shennan S. & Steele J. (2003) Neolithic transition in Europe: the radiocarbon record revisited. *Antiquity* **77**, 45–62.
- Kumar S., Tamura K., Jakobsen I.B. & Nei M. (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**, 1244–5.
- Larrea K., Napal S., Reta M. & Villanueva R. (2005) *Cuatro siglos de Casta Navarra (1605–2005). Pasado, presente y futuro*. Editorial Evidencia Médica S.L., Pamplona, Spain.
- Liron J.P., Bravi C.M., Mirol P.M., Peral-García P. & Giovambattista G. (2006) African matrilineages in American Creole cattle: evidence of two independent continental sources. *Animal Genetics* **37**, 379–82.
- Loftus R.T., MacHugh D.E., Bradley D.G., Sharp P.M. & Cunningham P. (1994) Evidence for two independent domestications of cattle. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 2757–61.
- Mannen H., Kohno M., Nagata Y. *et al.* (2004) Independent mitochondrial origin and historical genetic differentiation in North Eastern Asian cattle. *Molecular Phylogenetics and Evolution* **32**, 539–44.
- Miretti M.M., Pereira H.A. Jr, Poli M.A., Contel E.P. & Ferro J.A. (2002) African-derived mitochondria in South American native cattle breeds (*Bos taurus*): evidence of a new taurine mitochondrial lineage. *Journal of Heredity* **93**, 323–30.
- Miretti M.M., Dunner S., Naves M., Contel E.P. & Ferro J.A. (2004) Predominant African-derived mtDNA in Caribbean and Brazilian Creole cattle is also found in Spanish cattle (*Bos taurus*). *Journal of Heredity* **95**, 450–3.
- Robertson A. (1953) A numerical description of breed structure. *Journal of Agricultural Science* **43**, 334–6.
- Sambrook J., Fritsch E.F. & Maniatis T. (1989). *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA.
- Troy C.S., MacHugh D.E., Bailey J.F., Magee D.A., Loftus R.T., Cunningham P., Chamberlain A.T., Sykes B.C. & Bradley D.G. (2001) Genetic evidence for Near-Eastern origins of European cattle. *Nature* **410**, 1088–91.
- Vilà C., Seddon J. & Ellegren H. (2005) Genes of domestic mammals augmented by backcrossing with wild ancestors. *Trends in Genetics* **21**, 214–8.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Mitochondrial DNA D-loop sequence variation between the six major haplotypes (T, T1, T2, T3, T4 and T1*) and the six new haplotypes for the Lidia breed (1–6) and their frequencies (%).

Table S2 Names of the sampled lineages, number of sampled herds and number of samples from each lineage.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors.

